REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
27-02-2012	final	15-04-2009 to 30-11-2011

4. TITLE AND SUBTITLE

TOWARD AN INTEGRATED FRAMWORK FOR DATA-EFFICIENT PARAMETRIC ADAPTIVE DECTECTION

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-09-1-0310

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Hongbin Li

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Stevens Institute of Technology
Castle Point on Hudson
Hoboken, NJ 07030

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFOSR
875 N. Randolph St.
Suite 325
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-OSR-VA-TR-2012-0534

12. DISTRIBUTION / AVAILABILITY STATEMENT

Distribution A - Approved for Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The conjugate-gradient (CG) algorithm is investigated for reduced-rank STAP detection. A family of CG matched filter (CG-MF) is developed by using the k-th iteration of the CG in solving the Wiener-Hopf equation. The performance the CG-MF detectors is examined for two cases. The first involves an arbitrary covariance matrix. It is shown that each CG-MF detector 1) yields the highest output SINR and smallest MSE among all linear solutions in the Krylov subspace; and 2) is CFAR with non-decreasing detection probability as k increases. The second is a structured case frequently encountered in practice, where the covariance matrix contains a rank-r component due to dominant interference sources, a scaled identity due to the presence of white noise, and a perturbation component containing the residual interference and/or due to the estimation error. It is shown via a perturbation analysis that the (r+1)-st CG-MF detector achieves an output SINR nearly identical to that of the optimum MF detector which requires full iterations of the CG algorithm. Finally, the CG algorithm is used to solve a linear prediction problem in the parametric adaptive matched filter (PAMF). It is shown that the PAMF can be casted within the framework of reduced-rank STAP detection.

15. SUBJECT TERMS

Space-time adaptive processing (STAP), reduced-rank detection, signal-to-interference-and-noise ratio (SINR), constant false-alarm rate (CFAR), efficient algorithms

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**		47	Hongbin Li
					19b. TELEPHONE NUMBER *(include area code)* (201) 216-5604

Contents

Chapter 1

Executive Summary

The conjugate gradient (CG) algorithm is a popular method for solving a system of linear equations. It is computationally efficient and guaranteed to converge in a fixed number of iterations. As an iterative method, CG provides a series of approximations to the solution in an expanding Krylov subspace. In this project, the CG algorithm is explored for reduced-rank target detection in space-time adaptive processing (STAP). There are two specific goals. One is to develop reduced-rank STAP detectors based on the CG algorithm and investigate their detection performance. The other is to use to the CG algorithm to build a relation between the parametric adaptive matched filter (PAMF) and reduced-rank detection and develop efficient implementations for the PAMF detector.

Toward the development of CG-based reduced-rank detectors, the CG algorithm is used as an efficient solver to compute the weight vector of the matched filter (MF). As an iterative algorithm, it produces a series of approximations to the MF weight vector, each of which can be used to filter the test signal and form a test statistic. This effectively leads to a family of detectors, referred to as the CG-MF detectors, which are indexed by k the number of iterations incurred. We first consider a general case involving an arbitrary covariance matrix of the disturbance (including interference, noise, etc.) and show that all CG-MF detectors attain constant false alarm rate (CFAR) and, furthermore, are optimum in the sense that the k-th CG-MF detector yields the highest output signal-to-interference-and-noise ratio (SINR) among all linear detectors within the k-th Krylov subspace. We then consider a structured case frequently encountered in practice, where the covariance matrix of the disturbance contains a low-rank component (rank-r) due to dominant interference sources, a scaled identity due to the presence of a white noise, and a perturbation component containing the residual interference. We show that the $(r+1)$-st CG-MF detector achieves CFAR and an output SINR nearly identical to that of the MF detector which requires complete iterations of the CG algorithm till reaching convergence. Hence, the $(r+1)$-st CG-MF detector can be used in place of the MF detector for significant computational saving when r is small. Numerical results are presented to verify the accuracy of our analysis for the CG-MF detectors.

Originally, the PAMF detector was introduced by using a multichannel autoregressive (AR) parametric model for the disturbance signal in STAP detection. While the parametric approach

brings in benefits such as reduced training and computational requirements as compared with fully adaptive STAP detectors, the PAMF detector as a reduced-rank solution remains unclear. Toward the development of a reduced-rank detection framework for the PAMF detector, the CG algorithm to solve the linear prediction problem arising in the PAMF detector. It is shown that CG yields not only a new computationally efficient implementation of the PAMF detector, a new and efficient AR model order selection method that can naturally be integrated with CG iterations, but it also offers new perspectives of PAMF as a reduced-rank subspace detector. We first consider the integration of the CG algorithm with the MF and parametric matched filter (PMF) when the covariance matrix of the disturbance signal is known. It is then extended to the adaptive case where the covariance matrix is estimated from training data. Important issues such as computational complexity and convergence rate are discussed. Performance of the proposed CG-PAMF detector is examined by using the KASSPER and other computer generated data.

Chapter 2

Conjugate Gradient Matched Filter

2.1 Introduction

This project is concerned with the multichannel signal detection problem from temporally and spatially correlated clutter and/or jammer background, which is found in phased-array radar, sonar and many other applications. A widely explored technique for multichannel signal detection is space-time adaptive processing (STAP) [1], first proposed by Brennan, Reed and Mallett [2]. Most STAP-based methods, such as the adaptive matched filter (AMF) [3] and Kelly's generalized likelihood ratio test (GLRT) [4], need to invert a large space-time covariance matrix, thus requiring a substantial amount of secondary or training signals as well as a high computational cost. Aimed at mitigating the demanding training and computational requirements of the full-dimensional STAP methods, reduced-rank STAP techniques, such as eigencanceler [5], principal-component method [6], cross-spectral metric [7], multistage Wiener filter (MWF) [8], etc., have been proposed to reduce the dimension of the data in advance of detection. Such reduced dimensional techniques have been extensively studied; see, e.g., [9] and references therein.

The conjugate gradient (CG) algorithm (e.g,. [10]) is a popular method for solving a system of linear equations. It is computationally efficient and guaranteed to converge in a fixed number of iterations. As an iterative method, CG provides a series of approximations to the solution in an expanding Krylov subspace [10]. The CG algorithm has been explored for STAP detection and beamforming in a multitude of recent studies. Connections between the CG algorithm and the MWF for STAP detection are explored in [11, 12], and an equivalence is found between the CG and the multistage Wiener filters. For beamforming, a Krylov subspace adaptive beamformer is introduced to improve the robustness of model order determined in [13] and separate the desired signal from interference in [14]. In [15], the CG algorithm is employed to solve a linear prediction problem in the parametric adaptive matched filter (PAMF) [16, 17], which also shows that the PAMF detector is a member of the reduced-dimensional STAP family.

In this chapter, we explore the CG algorithm in the matched filter (MF) for STAP detection. The CG algorithm is employed to iteratively calculate the weight vector of the MF detector, which produces a series of progressively improved approximations to the MF weight vector. As we shall show, each of the intermediate weight vectors generated by CG iterations can be used to

form a decision variable, which, collectively, forms a family of STAP detectors referred to herein as the *CG-MF detectors*. Our goal is to examine the performance of these CG-MF detectors relative to the benchmark MF detector. The motivation is that if any detector within the CG-MF family offers a performance close to that of the MF detector, then the latter can be substituted by the former, especially if the CG-MF detector is obtained using a few iterations and, as such, is computational more efficient. We consider two cases, one involving a general covariance for the disturbance (i.e., clutter, jamming, and noise) and the other involving a structured disturbance covariance.

Specifically, for the first case, the space-time covariance matrix of the disturbance is assumed positive definite but otherwise arbitrary. The probability of false alarm and probability of detection of the CG-MF detectors are derived in closed form. The analysis shows that the CG-MF detectors achieve constant false alarm rate (CFAR), irrespective of the number of iterations. The probability of detection, for a given false alarm probability, is dictated by the output signal-to-interference-and-noise ratio (SINR) of the linear filter employed by the CG-MF detector, which changes over the CG iterations. It is found that the CG-MF detector obtained at the k-th CG iteration is optimum in the sense that it yields the largest output SINR over all linear detectors within the k-dimensional Krylov subspace. As a result, the output SINR and detection probability of the CG-MF detector improves with more CG iterations and converges to their counterparts of the MF detector.

For the second case, the disturbance space-time covariance matrix is assumed to include a low-rank component $\boldsymbol{R}_{\rm i}$, a scaled identity due to the presence of a white noise in the disturbance, and a perturbation component $\boldsymbol{\Delta}$ composed of entries that are generally small compared with those in $\boldsymbol{R}_{\rm i}$. The above structure is frequently encountered in many applications, whereby $\boldsymbol{R}_{\rm i}$ contains the dominant clutter scatterers or interference sources that need be effectively mitigated, whereas $\boldsymbol{\Delta}$ may include insignificant, residual clutter/interference sources. In the absence of the perturbation, it is well-known that the CG algorithm converges in $r+1$ iterations, where r is the rank of $\boldsymbol{R}_{\rm i}$ [10], and if r is small, requires significant less computation than directly computing the matrix inverse needed by the MF detector. However, with the presence of $\boldsymbol{\Delta}$, the CG algorithm in general requires full iterations, i.e., JN iterations with J being the number of channels and N the number of temporal samples, to reach convergence. Despite this, for the obvious reason of complexity reduction, there is an interest to investigate if we can abort the CG algorithm after $r+1$ iterations, and use the $(r+1)$-st CG-MF detector within the CG-MF family in place of the MF detector, provided that the perturbation is small. To answer the question, an analysis is provided which reveals a relation between the weight vectors of the $(r+1)$-st CG-MF detector and, respectively, the MF. The result is next used to compute and compare the output SINRs of the two detectors. Interestingly, it is found that the output SINRs are identical within a first-order approximation, that is, the output SINR loss experienced by the CG-MF detector relative to the MF is a higher order term of the perturbation which can be neglected for small perturbation. As such, the detection probabilities are also nearly identical, and be computed by using the results obtained for the first (general covariance matrix) case.

The remainder of this chapter is organized as follows. The signal detection problem is introduced in Section 2.2 and the CG-MF detector described in Section 2.3. The performance analysis

of the CG-MF detectors is presented in Section 2.4, including the general case of disturbance covariance matrix in Section 2.4.1 and the structured case in Section 2.4.2. Our analytical results are verified by computer simulations in Section 2.5. Finally, conclusions are summarized in Section 3.7.

Vectors and matrices are denoted by boldface lower-case and upper-case letters, respectively. Transpose, complex conjugate and complex conjugate transpose are respectively represented by $(\cdot)^T$, $(\cdot)^*$ and $(\cdot)^H$. $\mathbb{C}$ and $\mathbb{R}$ denote the complex and real number fields. $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{R})$ denotes a multivariate Gaussian random variable with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{R}$.

2.2 Data Model

Consider a J-channel sequence $\boldsymbol{x}(n) \in \mathbb{C}^{J\times 1}$, $n = 1, 2, \cdots, N$, received using an array antenna which is corrupted by a space-time correlated disturbance random process $\boldsymbol{c}(n)$. The detection problem involves the following binary hypotheses:

$$\begin{aligned} H_0 &: \ \boldsymbol{x}(n) = \boldsymbol{c}(n) \\ H_1 &: \ \boldsymbol{x}(n) = a\boldsymbol{s}(n) + \boldsymbol{c}(n) \end{aligned} \tag{2.1}$$

where $\boldsymbol{s}(n)$ is a known J-channel signal and a is its unknown complex amplitude. For the convenience of later discussions, define the following $JN\times 1$ vectors: $\boldsymbol{s} = [\boldsymbol{s}^T(1), \boldsymbol{s}^T(2), \cdots, \boldsymbol{s}^T(N)]^T$, $\boldsymbol{c} = [\boldsymbol{c}^T(1), \boldsymbol{c}^T(2), \cdots, \boldsymbol{c}^T(N)]^T$, and $\boldsymbol{x} = [\boldsymbol{x}^T(1), \boldsymbol{x}^T(2), \cdots, \boldsymbol{x}^T(N)]^T$. In STAP, $\boldsymbol{s}$ is known as the space-time steering vector. For a side-looking uniform linear array (ULA), $\boldsymbol{s}$ is given by

$$\boldsymbol{s} = \boldsymbol{s}_{\mathrm{t}} \otimes \boldsymbol{s}_{\mathrm{s}} \tag{2.2}$$

where $\boldsymbol{s}_{\mathrm{t}} = (1/\sqrt{N})\left[1, e^{i2\pi f_{\mathrm{d}}}, , \cdots, e^{i2\pi(N-1)f_{\mathrm{d}}}\right]^T$ is the temporal steering vector with a normalized Doppler frequency f_{d}, $\boldsymbol{s}_{\mathrm{s}} = (1/\sqrt{J})\left[1, e^{i2\pi f_{\mathrm{s}}}, , \cdots, e^{i2\pi(J-1)f_{\mathrm{s}}}\right]^T$ is the spatial steering vector with a normalized spatial frequency f_{s}, and $\otimes$ denotes the Kronecker product. For STAP detection, a standard assumption is that the disturbance $\boldsymbol{c}$ is a Gaussian random vector with zero-mean and space-time covariance matrix $\boldsymbol{R} \in \mathbb{C}^{JN\times JN}$ [1]. It follows that $\boldsymbol{x} \sim \mathcal{CN}(a\boldsymbol{s}, \boldsymbol{R})$, where $a = 0$ under H_0 and $a \neq 0$ under H_1.

2.3 Conjugate Gradient Matched Filter

The optimum detector for (2.1) is the matched filter (MF) (e.g., [3]):

$$t_{\mathrm{MF}} = \frac{|\boldsymbol{s}^H\boldsymbol{R}^{-1}\boldsymbol{x}|^2}{\boldsymbol{s}^H\boldsymbol{R}^{-1}\boldsymbol{s}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{2.3}$$

where η is the threshold of MF. The performance of the MF is regarded as a benchmark for all linear tests. For notational convenience, we will frequently denote a detector by a linear filter weight vector. For the MF detector, the weight vector is

$$\boldsymbol{w}_{\mathrm{MF}} = \boldsymbol{R}^{-1}\boldsymbol{s}. \tag{2.4}$$

Then the MF test statistic can be alternatively expressed as:

$$t_{\text{MF}} = \frac{|\boldsymbol{w}_{\text{MF}}^H \boldsymbol{x}|^2}{\boldsymbol{w}_{\text{MF}}^H \boldsymbol{R} \boldsymbol{w}_{\text{MF}}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \tag{2.5}$$

For typical STAP applications, the covariance matrix $\boldsymbol{R}$ has a large dimension. As a result, direct matrix inversion is usually not recommended to compute the weight vector (2.4) due to its computational complexity. We consider herein an alternative way by employing the conjugate (CG) algorithm [10], which iteratively finds a sequence of linear weight vectors $\boldsymbol{w}_k, k = 0, 1, \ldots$, that are guaranteed to converge to the MF weight vector in no more than JN iterations. Each of the weight vector $\boldsymbol{w}_k$ can be used to form a detector as in (2.5). As such, the CG iterations yield a family of detectors, referred to as the CG-MF detectors. The CG-MF detectors are closely related to the so-called multistage Wiener filter [11]. We examine the performance of these CG-MF detectors relative to the MF detector. To introduce necessary notation, The CG algorithm is briefly summarized as follows.

Initialization. Initialize the conjugate direction vector $\boldsymbol{d}_1$, gradient vector $\boldsymbol{\gamma}_1$ and initial solution $\boldsymbol{w}_0$:

$$\boldsymbol{d}_1 = -\boldsymbol{\gamma}_1 = \boldsymbol{s} \tag{2.6}$$

$$\boldsymbol{w}_0 = \boldsymbol{0}. \tag{2.7}$$

for $k = 1, 2, \cdots$, till convergence ($k \leq JN$) **do**

Update the step size α_k:

$$\alpha_k = \frac{\|\boldsymbol{\gamma}_k\|^2}{\boldsymbol{d}_k^H \boldsymbol{R} \boldsymbol{d}_k}. \tag{2.8}$$

Update the solution $\boldsymbol{w}_k$:

$$\boldsymbol{w}_k = \boldsymbol{w}_{k-1} + \alpha_k \boldsymbol{d}_k. \tag{2.9}$$

Update the gradient vector $\boldsymbol{\gamma}_{k+1}$:

$$\boldsymbol{\gamma}_{k+1} = \boldsymbol{\gamma}_k + \alpha_k \boldsymbol{R} \boldsymbol{d}_k. \tag{2.10}$$

Update the conjugate direction vector $\boldsymbol{d}_{k+1}$:

$$\boldsymbol{d}_{k+1} = \boldsymbol{d}_k \frac{\|\boldsymbol{\gamma}_{k+1}\|^2}{\|\boldsymbol{\gamma}_k\|^2} - \boldsymbol{\gamma}_{k+1}. \tag{2.11}$$

end for

A quick comment on the complexity is in order. Each iteration of the CG algorithm involves one matrix-vector product, requiring about $O((JN)^2)$ flops. With full JN iterations, the CG algorithm has a complexity of $O((JN)^3)$ flops, comparable with alternative linear solvers such as the QR factorization [10]. In many practical cases, the CG algorithm may require far fewer than full iterations (see Section 2.4.2 and also [10, Chap. 10] for discussions on the convergence of the CG algorithm), leading to significant reduction in complexity.

2.4 Analysis

In this section, we consider the performance of the CG-MF detectors under two cases. The first involves a general covariance matrix $\boldsymbol{R}$ that is positive definite but otherwise arbitrary, whereas the other deals with a structured covariance that is frequently encountered in practice.

2.4.1 General Covariance matrix

It would be useful to represent the CG-MF detector $\boldsymbol{w}_k$ by using the conjugate direction vectors $\{\boldsymbol{d}_k\}$ in closed form. First, (2.9) implies that

$$\boldsymbol{w}_k = \boldsymbol{D}_k \boldsymbol{\alpha}_k \tag{2.12}$$

where $\boldsymbol{\alpha}_k = [\alpha_1, \alpha_2, \cdots, \alpha_k]^T$ contains the stepsizes and $\boldsymbol{D}_k = [\boldsymbol{d}_1, \boldsymbol{d}_2, \cdots, \boldsymbol{d}_k]$ consists of the first k conjugate direction directors. One property of $\boldsymbol{D}_k$ is that it diagonalizes the covariance matrix $\boldsymbol{R}$ [10, p. 523]:

$$\boldsymbol{D}_k^H \boldsymbol{R} \boldsymbol{D}_k = \boldsymbol{\Lambda}_k \tag{2.13}$$

where $\boldsymbol{\Lambda}_k = \text{diag}(u_1^2, u_2^2, \cdots, u_k^2)$ and $u_k = (\boldsymbol{d}_k^H \boldsymbol{R} \boldsymbol{d}_k)^{\frac{1}{2}}$. This allows $\boldsymbol{\alpha}_k$ to be compactly expressed as

$$\boldsymbol{\alpha}_k = \boldsymbol{\Lambda}_k^{-1} \boldsymbol{D}_k^H \boldsymbol{s} \tag{2.14}$$

which gives the following close-form expression for $\boldsymbol{w}_k$:

$$\boldsymbol{w}_k = \boldsymbol{D}_k \boldsymbol{\Lambda}_k^{-1} \boldsymbol{D}_k^H \boldsymbol{s}. \tag{2.15}$$

The k-th CG-MF detector using $\boldsymbol{w}_k$ is given by

$$t_k = \frac{|\boldsymbol{w}_k^H \boldsymbol{x}|^2}{\boldsymbol{w}_k^H \boldsymbol{R} \boldsymbol{w}_k} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_k. \tag{2.16}$$

Theorem 1 *For the k-th CG-MF detector $\boldsymbol{w}_k$,*

(a) It is a linear minimum mean square estimator that minimizes

$$\boldsymbol{w}_k = arg \min_{\boldsymbol{w} \in \mathcal{K}(\boldsymbol{R}, \boldsymbol{s}, k)} \left\{ E\|a - \boldsymbol{w}^H \boldsymbol{x}\|^2 \right\} \tag{2.17}$$

among all linear estimators within the k-dimensional Krylov subspace

$$\mathcal{K}(\boldsymbol{R}, \boldsymbol{s}, k) \triangleq span\{\boldsymbol{s}, \boldsymbol{R}\boldsymbol{s}, \boldsymbol{R}^2\boldsymbol{s}, \cdots, \boldsymbol{R}^{k-1}\boldsymbol{s}\}.$$

(b) $\boldsymbol{w}_k$ yields the largest output SINR

$$\rho_k = |a|^2 \frac{|\boldsymbol{w}_k^H \boldsymbol{s}|^2}{\boldsymbol{w}_k^H \boldsymbol{R} \boldsymbol{w}_k} \tag{2.18}$$

among all linear detectors within $\mathcal{K}(\boldsymbol{R}, \boldsymbol{s}, k)$.

(c) Under the aforementioned Gaussian assumption, namely $\boldsymbol{x} \sim \mathcal{CN}(a\boldsymbol{s}, \boldsymbol{R})$ with $a = 0$ under H_0 and $a \neq 0$ under H_1, the probability of false alarm $P_{fa,k}$ and the probability of detection $P_{d,k}$ of $\boldsymbol{w}_k$ are

$$P_{fa,k} = \exp(-\eta_k) \tag{2.19}$$

$$P_{d,k} = Q(\sqrt{2\rho_k}, \sqrt{2\eta_k}) \tag{2.20}$$

where $Q(\cdot)$ is the Marcum Q function.

Sketch of proof: As the above results are quick extensions from standard knowledge, we only provide a sketch of proof. Result (a) is proved by observing that the Krylov subspace is also spanned by the conjugate direction vectors [10]: $\mathcal{K}(\boldsymbol{R}, \boldsymbol{s}, k) = \text{span}\{\boldsymbol{d}_1, \boldsymbol{d}_2, \cdots, \boldsymbol{d}_k\}$. As such,

$$\begin{aligned} \boldsymbol{w}_k =& \arg \min_{\boldsymbol{w}=\boldsymbol{D}_k\boldsymbol{\alpha}} \left\{\mathrm{E}\|a - \boldsymbol{w}^H\boldsymbol{x}\|^2\right\} \\ =& \boldsymbol{D}_k \left(\boldsymbol{D}_k^H \boldsymbol{R} \boldsymbol{D}_k\right)^{-1} \boldsymbol{D}_k^H \boldsymbol{s} = \boldsymbol{D}_k \boldsymbol{\Lambda}_k^{-1} \boldsymbol{D}_k^H \boldsymbol{s} \end{aligned} \tag{2.21}$$

which is identical to (2.15).

Result (b) follows immediately from (a), since minimizing the mean-square error (MSE) is equivalent to maximizing the SINR [18].

To show (c), we use (2.13) and (2.12) to rewrite the test variable t_k of (2.16) as

$$t_k = \frac{|\boldsymbol{\alpha}_k^H \boldsymbol{D}_k^H \boldsymbol{x}|^2}{\boldsymbol{\alpha}_k^H \boldsymbol{\Lambda}_k \boldsymbol{\alpha}_k} \boldsymbol{\alpha}_k^H \boldsymbol{\Lambda}_k \boldsymbol{\alpha}_k = |y_k|^2 \tag{2.22}$$

where $y_k = (\boldsymbol{\alpha}_k^H \boldsymbol{D}_k^H \boldsymbol{x})(\boldsymbol{\alpha}_k^H \boldsymbol{\Lambda}_k \boldsymbol{\alpha}_k)^{-\frac{1}{2}}$. It is clear that $y_k \sim \mathcal{CN}(0, 1)$ under H_0 and, respectively, $y_k \sim \mathcal{CN}\left((a\boldsymbol{\alpha}_k^H \boldsymbol{D}_k^H \boldsymbol{s})(\boldsymbol{\alpha}_k^H \Lambda_k \boldsymbol{\alpha}_k)^{-\frac{1}{2}}, 1\right)$ under H_1. As such, t_k is central and, respectively, noncentral Chi-square distributed under H_0 and H_1, from which (2.19) and (2.20) follow immediately. Q.E.D.

A number of remarks are in order. First, (2.19) implies that CG-MF detectors for all k are CFAR detectors. Their test variables t_k are all identically distributed to that of the MF test variable t_{MF}, irrespective of k the number of iterations. Second, (b) implies that $\rho_k \leq \rho_{k+1}$, since $\mathcal{K}(\boldsymbol{R}, \boldsymbol{s}, k) \subseteq \mathcal{K}(\boldsymbol{R}, \boldsymbol{s}, k+1)$. Hence, the CG-MF is a family of CFAR detectors $\boldsymbol{w}_k$ composed of both reduced-rank detectors ($k < JN$) and the full-rank MF detector ($k = JN$), which offers a natural way to trade complexity for performance. Specifically, the detection probability $P_{\text{d},k}$ of the CG-MF detector $\boldsymbol{w}_k$ increases with more CG iterations (i.e., a larger k), at higher computational complexity. The trade-off and the analytical expression (2.20) allow one to save the computational cost by selecting an appropriate reduced-rank CG-MF detector that offers a targeted $P_{\text{d},k}$, without going through all CG iterations.

2.4.2 CG-MF: Structured Covariance Matrix

One desired property of the CG algorithm is its fast convergence. In general, it takes no more than JN iterations to obtain the MF detector vector $\boldsymbol{w}_{\text{MF}}$ (2.4) [10]. Even faster convergence

is possible if the covariance matrix of the disturbance has some specific structure, which makes the CG algorithm more attractive than many alternative approaches to computing (2.4). In this section, we examine the performance of the CG-MF detectors in the case where the disturbance covariance matrix has some low-rank structure. Specifically, suppose $\boldsymbol{R}$ consists of the following two components:

$$\boldsymbol{R}_{\text{i}} + \sigma_{\text{n}}^2 \boldsymbol{I} \tag{2.23}$$

where $\boldsymbol{R}_{\text{i}}$ is a rank-r positive semi-definite matrix ($r < JN$) and $\boldsymbol{I}$ an identify matrix. For such a structured matrix, it is known that the CG algorithm converges to the baseline MF solution using at most $r+1$ iterations [10], i.e., $\boldsymbol{w}_{r+1} = \boldsymbol{w}_{\text{MF}}$, which is computationally very efficient.

Many practical applications involve a disturbance covariance matrix having a structure similar to (2.23). For example, in airborne radar applications, the disturbance covariance matrix often consists of two components, namely a low-rank $\boldsymbol{R}_{\text{i}}$ due to the clutter and jamming and a scaled identity $\sigma_{\text{n}}^2 \boldsymbol{I}$ due to the thermal noise, where σ_{n}^2 denotes the noise variance. The rank r is typically much smaller than the joint spatio-temporal dimension JN. Specifically, if the disturbance is primarily due to ground clutter and thermal noise, then according to Brennan's rule [2], the rank of the clutter covariance matrix for the full-dimensional MF is approximately

$$r \approx \lceil J + (N-1)\beta \rceil \tag{2.24}$$

where $\beta = 2v_{\text{g}}T_{\text{r}}/d$, v_{g} is the platform velocity, T_{r} is the pulse repetition period, d is the antenna element spacing, and $\lceil \cdot \rceil$ rounds a real-valued number towards infinity.

Estimating the rank r can be a tricky issue since (2.24) may not hold for all clutter scenarios encountered in practice. The CG algorithm has an advantage of not requiring to know r a priori, since at the $(r+1)$-st iteration, the residual $\boldsymbol{s} - \boldsymbol{R}\boldsymbol{w}_k$, which is also the negative gradient $\boldsymbol{\gamma}_k$, vanishes. This is the stopping rule used by the CG algorithm to stop iterations [10]. Other STAP detectors designed to take advantage of the structure (2.23), such as the low rank normalized matched filter (LRNMF) [19] which employs the principal eigenvectors of the covariance matrix, requires an estimate of r and its performance is quite sensitive to the accuracy of the estimate.

While the convergence of the CG for a structured covariance matrix exactly like (2.23) is well known, in the following, we consider a different but related scenario that the disturbance covariance matrix $\boldsymbol{R}$ is a perturbed version of (2.23):

$$\boldsymbol{R} = \boldsymbol{R}_{\text{i}} + \sigma_{\text{n}}^2 \boldsymbol{I} + \boldsymbol{\Delta} \triangleq \boldsymbol{R}_0 + \boldsymbol{\Delta} \tag{2.25}$$

where $\boldsymbol{R}_0 = \boldsymbol{R}_{\text{i}} + \sigma_{\text{n}}^2 \boldsymbol{I}$ is structured as in (2.23) and $\boldsymbol{\Delta}$ is a Hermitian perturbation matrix that is assumed to be small, i.e., $\|\boldsymbol{\Delta}\| \ll \|\boldsymbol{R}_0\|$. Since the perturbation is small, it is of interest to examine the following important questions: Can we benefit from the CG algorithm when the disturbance covariance matrix is given as (2.25)? Can it reach convergence or almost convergence in $r+1$ iterations? How is the detection performance of the CG-MF detector $\boldsymbol{w}_{r+1}$ compared to the MF detector? A second type of perturbation is stochastic perturbation arising from finite training data effects in estimating the covariance matrix. The random perturbation issue is a subject of ongoing research and will be reported in a future publication.

Before we address the above questions, some discussions on the model (2.25) are in order. The perturbation model may exist in many scenarios. For example, in airborne radar applications,

the covariance matrix $\boldsymbol{R}$ rarely has exactly $r+1$ distinct eigenvalues as in (2.23). Typically, $\boldsymbol{R}$ contains a few principal eigenvalues due to the dominant clutter scatterers, but the other eigenvalues are rarely identical and spread around the noise level [1]. By decomposing $\boldsymbol{R}$ as (2.25), $\boldsymbol{R}_{\rm i}$ contain only the dominant clutter scatterers, and the effect of the less significant clutter scatterers can be included in $\boldsymbol{\Delta}$. The same can be extended to a general interference scenario, where $\boldsymbol{R}_{\rm i}$ is the covariance matrix of only a few major interference sources that have to be mitigated at the receiver, whereas $\boldsymbol{\Delta}$ contains the residual interference.

Given a structured covariance matrix $\boldsymbol{R}$ with perturbation as in (2.25), the analytical results of Section 2.4.1 of the family of CG-MF detectors are still applicable. To address the earlier questions, we consider the CG-MF detector $\boldsymbol{w}_{r+1}$ obtained at the $(r+1)$-st CG iteration, and examine the performance of this CG-MF detector relative to the MF detector $\boldsymbol{w}_{\rm MF}$. We first present a result that relates the weight vectors for the two detectors.

Lemma 1 *Consider the linear equation $\boldsymbol{R}\boldsymbol{w}_{MF}=\boldsymbol{s}$, where $\boldsymbol{R}=\boldsymbol{R}_{\rm i}+\sigma_{\rm n}^2\mathbf{I}+\boldsymbol{\Delta}=\boldsymbol{R}_0+\boldsymbol{\Delta}$ is a positive-definite Hermitian matrix, $\boldsymbol{R}_{\rm i}$ is a rank-r positive semi-definite Hermitian matrix, $\sigma_n^2>0$ is a constant, and $\boldsymbol{\Delta}$ is a Hermitian perturbation matrix. If the perturbation is small such that $\|\boldsymbol{\Delta}\|\ll\|\boldsymbol{R}_0\|$, the MF solution $\boldsymbol{w}_{MF}$ can be approximated by the CG-MF solution $\boldsymbol{w}_{r+1}$, with the approximation error given by*

$$\boldsymbol{w}_{MF}-\boldsymbol{w}_{r+1}=\boldsymbol{R}_0^{-\frac{1}{2}}\boldsymbol{P}_{\tilde{\boldsymbol{T}}_r}^{\perp}\boldsymbol{R}_0^{-\frac{1}{2}}\boldsymbol{d}+o(\|\boldsymbol{\Delta}\|) \tag{2.26}$$

where $o(\|\boldsymbol{\Delta}\|)$ contains the second- and higher-order perturbation terms that can be neglected for small $\|\boldsymbol{\Delta}\|$,

$$\tilde{\boldsymbol{T}}_r=\boldsymbol{R}_0^{\frac{1}{2}}\boldsymbol{T}_r \tag{2.27}$$

$$\boldsymbol{T}_r=[\boldsymbol{s},\boldsymbol{R}_0\boldsymbol{s},\boldsymbol{R}_0^2\boldsymbol{s},\cdots,\boldsymbol{R}_0^r\boldsymbol{s}] \tag{2.28}$$

$$\boldsymbol{P}_{\tilde{\boldsymbol{T}}_r}^{\perp}=\boldsymbol{I}-\tilde{\boldsymbol{T}}_r(\tilde{\boldsymbol{T}}_r^H\tilde{\boldsymbol{T}}_r)^{-1}\tilde{\boldsymbol{T}}_r^H \tag{2.29}$$

$$\boldsymbol{d}=\left(\boldsymbol{\Delta}\boldsymbol{R}_0^{-1}+\boldsymbol{R}_0\boldsymbol{\Phi}_{\Delta}(\boldsymbol{T}_r^H\boldsymbol{R}_0\boldsymbol{T}_r)^{-1}\boldsymbol{T}_r^H\right)\boldsymbol{s} \tag{2.30}$$

and

$$\boldsymbol{\Phi}_{\Delta}=\left[\mathbf{0},\boldsymbol{\Delta}\boldsymbol{s},\boldsymbol{R}_0\boldsymbol{\Delta}\boldsymbol{s}+\boldsymbol{\Delta}\boldsymbol{R}_0\boldsymbol{s},\cdots,\sum_{k=1}^{r}\boldsymbol{R}_0^{r-k}\boldsymbol{\Delta}\boldsymbol{R}_0^{k-1}\boldsymbol{s}\right]. \tag{2.31}$$

Proof: See Appendix 2.7. Q.E.D.

When the perturbation vanishes, it is straightforward to show from (2.51) of Appendix 2.7 that

$$\boldsymbol{w}_{\rm MF}-\boldsymbol{w}_{r+1}=\boldsymbol{R}_0^{-\frac{1}{2}}\boldsymbol{P}_{\tilde{\boldsymbol{T}}_r}^{\perp}\boldsymbol{R}_0^{-\frac{1}{2}}\boldsymbol{s}=\mathbf{0} \tag{2.32}$$

which offers another proof that the CG-MF detector $\boldsymbol{w}_{r+1}$ converges to the MF detector $\boldsymbol{w}_{\rm MF}$ when $\boldsymbol{\Delta}=\mathbf{0}$, a result we already know for the CG algorithm [10]. Interestingly, the first equality

of the above equation resembles (2.26) except that $\boldsymbol{d}$ is replaced by $\boldsymbol{s}$. It is also noted that the matrix $\boldsymbol{\Phi}_\Delta$ can be easily calculated by the following simple recursion:

$$\begin{aligned}\boldsymbol{\phi}_1 =& \mathbf{0}\\ \boldsymbol{\phi}_k =& \boldsymbol{\Delta}\boldsymbol{R}_0^{k-2}\boldsymbol{s} + \boldsymbol{R}_0\boldsymbol{\phi}_{k-1}, \quad k = 2, 3, \cdots, r+1.\end{aligned} \tag{2.33}$$

We now consider the output SINR of the CG-MF detector $\boldsymbol{w}_{r+1}$,

$$\rho_{r+1} = |a|^2 \frac{|\boldsymbol{w}_{r+1}^H \boldsymbol{s}|^2}{\boldsymbol{w}_{r+1}^H \boldsymbol{R}\boldsymbol{w}_{r+1}} \tag{2.34}$$

and its relation to the output SINR of the MF detector. The following result addresses their relationship.

Theorem 2 *Under the same conditions stated in Lemma 1, the output SINRs of the MF detector $\boldsymbol{w}_{\rm MF}$ and the CG-MF detector $\boldsymbol{w}_{r+1}$ are identical within a first-order approximation. That is,*

$$\delta_\rho = \rho_{\rm MF} - \rho_{r+1} = o(\|\boldsymbol{\Delta}\|) \tag{2.35}$$

where

$$\rho_{r+1} = |a|^2 \frac{|\boldsymbol{w}_{r+1}^H \boldsymbol{s}|^2}{\boldsymbol{w}_{r+1}^H \boldsymbol{R}\boldsymbol{w}_{r+1}}. \tag{2.36}$$

and $\rho_{\rm MF}$ is similarly defined by replacing $\boldsymbol{w}_{r+1}$ with $\boldsymbol{w}_{\rm MF}$.

Proof: The proof goes by direct calculation and using Lemma 1. The loss of output SINR of the CG-MF relative to the MF is given by

$$\delta_\rho = \rho_{\rm MF} - \rho_{r+1} = |a|^2 \left(\frac{|\boldsymbol{w}_{\rm MF}^H \boldsymbol{s}|^2}{\boldsymbol{w}_{\rm MF}^H \boldsymbol{R}\boldsymbol{w}_{\rm MF}} - \frac{|\boldsymbol{w}_{r+1}^H \boldsymbol{s}|^2}{\boldsymbol{w}_{r+1}^H \boldsymbol{R}\boldsymbol{w}_{r+1}} \right). \tag{2.37}$$

First we consider the difference between $\boldsymbol{s}^H\boldsymbol{w}_{\rm MF}$ and $\boldsymbol{s}^H\boldsymbol{w}_{r+1}$. Using Lemma 1, we have

$$\boldsymbol{s}^H\boldsymbol{w}_{\rm MF} - \boldsymbol{s}^H\boldsymbol{w}_{r+1} = \boldsymbol{s}^H \boldsymbol{R}_0^{-\frac{1}{2}} \boldsymbol{P}_{\tilde{\boldsymbol{T}}_r}^{\perp} \boldsymbol{R}_0^{-\frac{1}{2}} \boldsymbol{d} + o(\|\boldsymbol{\Delta}\|). \tag{2.38}$$

Since $\boldsymbol{s}$ is orthogonal to the column space of $\boldsymbol{R}_0^{-\frac{1}{2}} \boldsymbol{P}_{\tilde{\boldsymbol{T}}_r}^{\perp} \boldsymbol{R}_0^{-\frac{1}{2}}$ [also see (2.32)], we have

$$\boldsymbol{s}^H \boldsymbol{R}_0^{-\frac{1}{2}} \boldsymbol{P}_{\tilde{\boldsymbol{T}}_r}^{\perp} \boldsymbol{R}_0^{-\frac{1}{2}} = 0 \tag{2.39}$$

and (2.38) reduces to

$$\boldsymbol{s}^H\boldsymbol{w}_{\rm MF} - \boldsymbol{s}^H\boldsymbol{w}_{r+1} = o(\|\boldsymbol{\Delta}\|). \tag{2.40}$$

It follows that

$$|\boldsymbol{w}_{\rm MF}^H \boldsymbol{s}|^2 - |\boldsymbol{w}_{r+1}^H \boldsymbol{s}|^2 = o(\|\boldsymbol{\Delta}\|). \tag{2.41}$$

Next, we consider the difference between the denominators $\boldsymbol{w}_{\text{MF}}^H \boldsymbol{R} \boldsymbol{w}_{\text{MF}}$ and $\boldsymbol{w}_{r+1}^H \boldsymbol{R} \boldsymbol{w}_{r+1}$:

$$\begin{aligned}
&\boldsymbol{w}_{\text{MF}}^H \boldsymbol{R} \boldsymbol{w}_{\text{MF}} - \boldsymbol{w}_{r+1}^H \boldsymbol{R} \boldsymbol{w}_{r+1} \\
&= 2\mathfrak{Re}\{\boldsymbol{w}_{\text{MF}}^H \boldsymbol{R} (\boldsymbol{w}_{\text{MF}} - \boldsymbol{w}_{r+1})\} \\
&\quad - (\boldsymbol{w}_{\text{MF}} - \boldsymbol{w}_{r+1})^H \boldsymbol{R} (\boldsymbol{w}_{\text{MF}} - \boldsymbol{w}_{r+1}) \\
&= 2\mathfrak{Re}\{\boldsymbol{s}^H (\boldsymbol{w} - \boldsymbol{w}_{r+1})\} \\
&\quad - (\boldsymbol{w}_{\text{MF}} - \boldsymbol{w}_{r+1})^H \boldsymbol{R} (\boldsymbol{w}_{\text{MF}} - \boldsymbol{w}_{r+1})
\end{aligned} \tag{2.42}$$

where $\mathfrak{Re}\{\cdot\}$ denotes the real part. Again using Lemma 1, we have

$$(\boldsymbol{w}_{\text{MF}} - \boldsymbol{w}_{r+1})^H \boldsymbol{R} (\boldsymbol{w}_{\text{MF}} - \boldsymbol{w}_{r+1}) = o(\|\boldsymbol{\Delta}\|). \tag{2.43}$$

Substituting (2.40) and (2.43) into (2.42) yields

$$\boldsymbol{w}_{\text{MF}}^H \boldsymbol{R} \boldsymbol{w}_{\text{MF}} - \boldsymbol{w}_{r+1}^H \boldsymbol{R} \boldsymbol{w}_{r+1} = o(\|\boldsymbol{\Delta}\|). \tag{2.44}$$

Finally, from (2.37), the output SINR loss of the CG-MF detector is given by

$$\delta_\rho = |a|^2 \frac{|\boldsymbol{w}_{\text{MF}}^H \boldsymbol{s}|^2 \boldsymbol{w}_{r+1}^H \boldsymbol{R} \boldsymbol{w}_{r+1} - |\boldsymbol{w}_{r+1}^H \boldsymbol{s}|^2 \boldsymbol{w}_{\text{MF}}^H \boldsymbol{R} \boldsymbol{w}_{\text{MF}}}{\boldsymbol{w}_{\text{MF}}^H \boldsymbol{R} \boldsymbol{w}_{\text{MF}} \boldsymbol{w}_{r+1}^H \boldsymbol{R} \boldsymbol{w}_{r+1}}. \tag{2.45}$$

Substituting (2.41) and (2.44) into the numerator of (2.45), we have

$$\begin{aligned}
&|\boldsymbol{w}_{\text{MF}}^H \boldsymbol{s}|^2 \boldsymbol{w}_{r+1}^H \boldsymbol{R} \boldsymbol{w}_{r+1} - |\boldsymbol{w}_{r+1}^H \boldsymbol{s}|^2 \boldsymbol{w}_{\text{MF}}^H \boldsymbol{R} \boldsymbol{w}_{\text{MF}} \\
&= |\boldsymbol{w}_{\text{MF}}^H \boldsymbol{s}|^2 \left(\boldsymbol{w}_{\text{MF}}^H \boldsymbol{R} \boldsymbol{w}_{\text{MF}} - o(\|\boldsymbol{\Delta}\|) \right) \\
&\quad - \left(|\boldsymbol{w}_{\text{MF}}^H \boldsymbol{s}|^2 - o(\|\boldsymbol{\Delta}\|) \right) \boldsymbol{w}_{\text{MF}}^H \boldsymbol{R} \boldsymbol{w}_{\text{MF}} \\
&= o(\|\boldsymbol{\Delta}\|)
\end{aligned} \tag{2.46}$$

from which (2.35) immediately follows. Q.E.D.

Remark: It is interesting to note that while Lemma 1 indicates that the difference between the weight vectors, i.e., $\boldsymbol{w}_{\text{MF}} - \boldsymbol{w}_{r+1}$, contains first-order terms of the perturbation, such first-order differences vanish in terms of the output SINR, as shown in the above proof. Theorem 2 implies that that the probabilities of detection of the MF and CG-MF detectors are also identical within a first-order approximation. This has important practical implication. In particular, even though the CG algorithm using $\boldsymbol{R}$ generally requires full (i.e., JN) iterations before it reaches convergence (since $\boldsymbol{R}$ does not have the low-rank structure (2.23)), we can take the intermediate result $\boldsymbol{w}_{r+1}$ obtained at the $(r+1)$-st iteration and obtain nearly the same detection performance as the MF detector, provided that $\boldsymbol{\Delta}$ is sufficiently small. While the latter condition cannot be quantitatively specified, numerical simulation can be used to examine the effect of the perturbation size and the accuracy of the approximation.

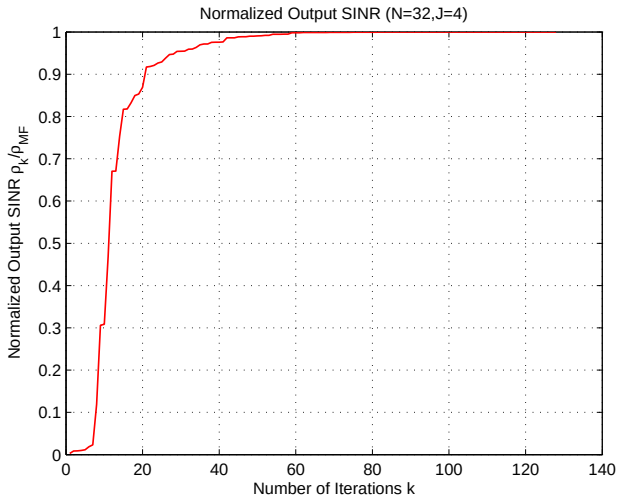

Figure 2.1: Normalized output SINR versus the number of iterations for the CG-MF detector ($J = 4$; $N = 32$).

2.5 Numerical Results

2.5.1 General Covariance Matrix

Computer simulation is employed to verify the analytical results reported in the previous sections. We first consider the general covariance matrix case studied in Section 2.4.1. Our numerical results for this case use a disturbance covariance matrix $\boldsymbol{R}$ obtained from the KASSPER data set [20], which is a simulated data set that includes practical airborne radar parameters and issues found in a real-world clutter environment. The radar platform considered in this data set has 11 horizontal antenna elements. For simplicity, we use only the outputs of the first $J = 4$ channels for processing. The number of pulses is $N = 32$, and the probability of false alarm is $P_{\text{fa}} = 0.01$. We first examine the output SINR ρ_k, defined in (2.18), of the CG-MF detector $\boldsymbol{w}_k$. Fig. 2.1 shows the normalized output SINR ρ_k/ρ_{MF}, where the normalizing factor ρ_{MF} is the output SINR of the MF detector, versus the number of iterations. It is observed that ρ_k converges rapidly to ρ_{MF} as k increases.

The probability of detection for the MF detector and the CG-MF detector after $k = 10, 20$ and 40 iterations, respectively, is shown in Fig. 2.2 as a function of the MF output SINR, defined as $\rho_{\text{MF}} \triangleq |a|^2 \boldsymbol{s}^H \boldsymbol{R}^{-1} \boldsymbol{s}$. It is seen that that with $k = 40$ iterations, the CG-MF detector achieves nearly identical detection performance as the MF detector, which requires $JN = 128$ CG iterations. Finally, the detection probability of the CG-MF detector as a function of the number of iterations k is shown in Fig. 2.3 for the MF output SINR $= 5, 10$, and 20 dB, respectively. It is seen that the detection probability of the CG-MF detector is a non-decreasing function of k in all cases, as predicted in Section 2.4.1.

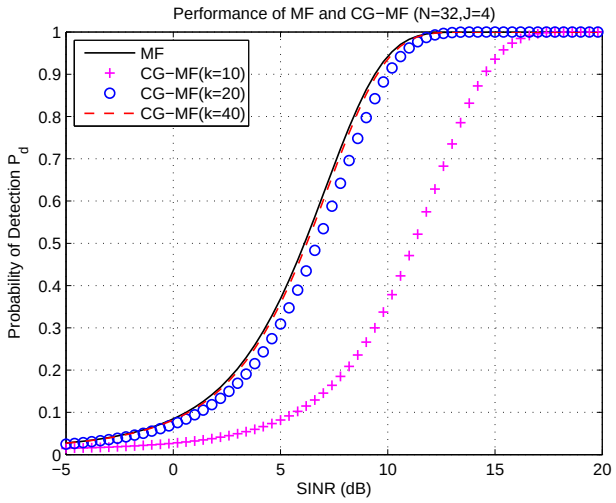

Figure 2.2: Probability of detection versus SINR for the MF and CG-MF detectors with several different numbers of iterations ($J = 4$; $N = 32$; $P_{\text{fa}} = 0.01$).

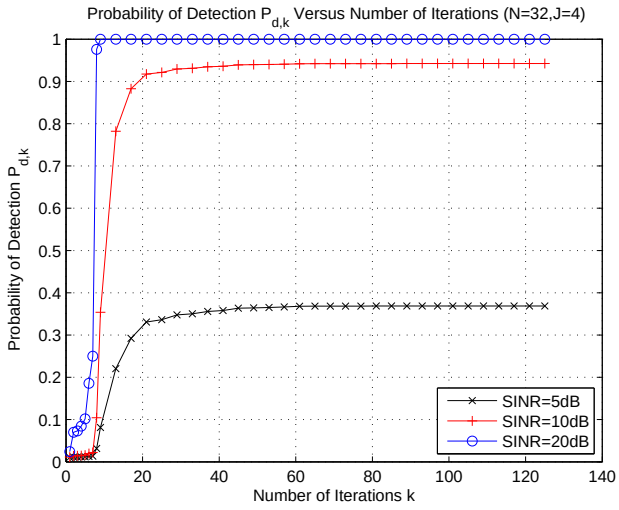

Figure 2.3: Probability of detection $P_{d,k}$ versus number of iterations k ($J = 4$; $N = 32$; $P_{\text{fa}} = 0.01$).

2.5.2 Structured Covariance Matrix with Perturbation

We now consider the case examined in Section 2.4.2 where the covariance $\boldsymbol{R}$ is a perturbed version of a structured $\boldsymbol{R}_0$. We demonstrate how the convergence of the CG-MF detector is directly affected by the size of the perturbation. We employ a *relative perturbation size*, defined as

$$a_p = \sqrt{\frac{\text{tr}(\boldsymbol{\Delta}^H \boldsymbol{\Delta})}{\text{tr}(\boldsymbol{R}_0^H \boldsymbol{R}_0)}} \tag{2.47}$$

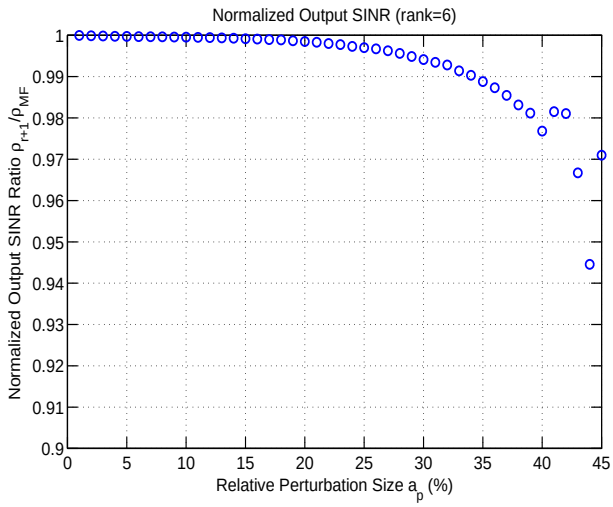

Figure 2.4: Output SINR of the MF detector and CG-MF detector after $r+1 = 7$ iterations versus the relative perturbation size ($r = 6$, $JN = 16$).

where tr$(\cdot)$ is the trace of a matrix. We tried several ways of generating the perturbation matrix $\boldsymbol{\Delta}$ and obtained similar results. The ones presented here were based on the following approach. For any structured covariance matrix $\boldsymbol{R}_0$ as described in (2.25) and a given perturbation size κ, 1) randomly generate $\boldsymbol{R}$ as a complex Wishhart matrix $\boldsymbol{R}'$ with mean $\boldsymbol{R}_0$; and 2) form $\boldsymbol{R}$ as a scaled version of $\boldsymbol{R}'$ such that $\boldsymbol{\Delta} = \boldsymbol{R} - \boldsymbol{R}_0$ has a prescribed perturbation size κ. It is noted that although $\boldsymbol{R}$ is generated as a random matrix whose mean is not $\boldsymbol{R}_0$ (which is the mean of $\boldsymbol{R}'$), in each trail $\boldsymbol{R}$ is treated as a deterministic/known matrix that is a perturbed version of $\boldsymbol{R}_0$ with perturbation size κ.

Fig. 2.4 shows the output SINR of the CG-MF detector $\boldsymbol{w}_{r+1}$ ($r = 6$) normalized by that of the MF detector, i.e., $\rho_{r+1}/\rho_{\text{MF}}$, as a function of a_p. It is seen that the output SINRs of the two detectors remain nearly identical ($\rho_{r+1}/\rho_{\text{MF}} > 0.99$) for a relative perturbation size as large as $a_p = 30\%$, which indicates that our perturbation analysis in Theorem 2 for the CG-MF detectors is quite accurate over a wide range of perturbation size. Fig. 2.5 depicts the probability of detection for the MF and CG-MF detector as a function of the MF output SINR, where several values of a_p are considered. It is seen that with a relative perturbation size as large as $a_p = 30\%$, the detection probability of the two detectors are nearly identical. At $a_p = 45\%$, a small difference is observed.

2.6 Concluding Remarks

The CG algorithm can be used to solve the Wiener-Hopf equation underlying the MF, which leads to a family of linear CG-MF detectors that converge to the MF in a fixed number of iterations. We have shown that the CG-MF detectors are all CFAR detectors, they can be recursively and efficiently computed via CG iterations over an expanding Krylov subspace, and each of them is an

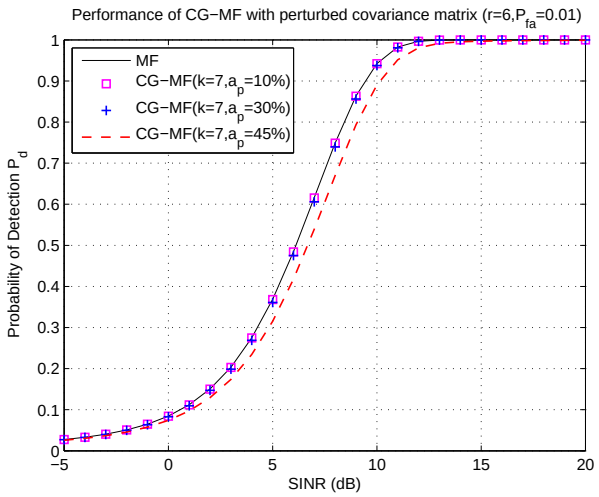

Figure 2.5: Probability of detection of the MF detector and the CG-MF detector after $r+1=7$ iterations ($r=6$, $JN=16$, $P_{\text{fa}}=0.01$).

optimum reduced-dimensional detector in the sense that it yields the maximum output SINR over all linear detectors residing the Krylov subspace. For disturbance covariances with a low-rank structure (rank-r), we have shown that the presence of a perturbation component $\boldsymbol{\Delta}$ disrupting the low-rank structure has minimum effect on the convergence of the CG algorithm, in that the output SINR of the $(r+1)$-st CG-MF detector is nearly identical to that of the MF detector. This offers significant computational saving, in particular when r is small, by using the CG-MF instead of the MF detector without incurring undue penalty in detection performance. A future topic of interest is to analyze the CG algorithm for adaptive detection when the covariance matrix $\boldsymbol{R}$ is unknown and estimated from training signals.

2.7 Appendix: Proof of Lemma 1

Proof: The CG-MF solution $\boldsymbol{w}_{r+1}$ obtained at the $(r+1)$-st iteration is the $\boldsymbol{R}$-orthogonal projection of $\boldsymbol{w}_{\text{MF}}$ onto the Krylov subspace $\mathcal{K}(\boldsymbol{R},\boldsymbol{s},r+1)$ [10]. This means that the $\boldsymbol{R}$-norm of the approximation error is minimized over all vectors in $\mathcal{K}(\boldsymbol{R},\boldsymbol{s},r+1)$, which is the column space of $\boldsymbol{S}_r=\left[\boldsymbol{s},\boldsymbol{R}\boldsymbol{s},\boldsymbol{R}^2\boldsymbol{s},\boldsymbol{R}^3\boldsymbol{s},\cdots,\boldsymbol{R}^r\boldsymbol{s}\right]$ [10]. That is,

$$\|\boldsymbol{w}_{\text{MF}}-\boldsymbol{w}_{r+1}\|_{\boldsymbol{R}}=\min_{a_k}\left\|\boldsymbol{w}-\sum_{k=0}^{r}a_k\boldsymbol{R}^k\boldsymbol{s}\right\|_{\boldsymbol{R}}. \tag{2.48}$$

Substituting $\boldsymbol{w} = \boldsymbol{R}^{-1}\boldsymbol{s}$ and $\|\cdot\|_{\boldsymbol{R}} = \left\|\boldsymbol{R}^{\frac{1}{2}}(\cdot)\right\|$ into (2.48), we have

$$\begin{aligned}\|\boldsymbol{w}_{\text{MF}} - \boldsymbol{w}_{r+1}\|_{\boldsymbol{R}} &= \min_{a_k} \left\|\boldsymbol{R}^{\frac{1}{2}}(\boldsymbol{R}^{-1}\boldsymbol{s} - \sum_{k=0}^{r} a_k \boldsymbol{R}^k \boldsymbol{s})\right\| \\ &= \min_{a_k} \left\|\boldsymbol{R}^{-\frac{1}{2}}\boldsymbol{s} - \sum_{k=0}^{r} a_k \boldsymbol{R}^{\frac{1}{2}}\boldsymbol{R}^k \boldsymbol{s}\right\|. \end{aligned} \tag{2.49}$$

The minimum approximation error is achieved if and only if the vector $\sum_{k=0}^{r} a_k \boldsymbol{R}^{\frac{1}{2}}\boldsymbol{R}^k\boldsymbol{s}$ is the orthogonal projection of the vector $\boldsymbol{R}^{-\frac{1}{2}}\boldsymbol{s}$ onto the linearly transformed Krylov subspace

$$\boldsymbol{R}^{\frac{1}{2}}\mathcal{K}(\boldsymbol{R}, \boldsymbol{s}, r+1) = \text{span}\{\boldsymbol{R}^{\frac{1}{2}}\boldsymbol{s}, \boldsymbol{R}^{\frac{1}{2}}\boldsymbol{R}\boldsymbol{s}, \cdots, \boldsymbol{R}^{\frac{1}{2}}\boldsymbol{R}^r\boldsymbol{s}\} \tag{2.50}$$

or the column space of $\tilde{\boldsymbol{S}}_r = \boldsymbol{R}^{\frac{1}{2}}\boldsymbol{S}_r$. When the minimum of (2.49) is achieved, the approximation error is given by

$$\boldsymbol{w}_{\text{MF}} - \boldsymbol{w}_{r+1} = \boldsymbol{R}^{-\frac{1}{2}}\boldsymbol{P}^{\perp}_{\tilde{\boldsymbol{S}}_r}\boldsymbol{R}^{-\frac{1}{2}}\boldsymbol{s} \tag{2.51}$$

where

$$\begin{aligned}\boldsymbol{P}^{\perp}_{\tilde{\boldsymbol{S}}_r} &= \boldsymbol{I} - \tilde{\boldsymbol{S}}_r(\tilde{\boldsymbol{S}}_r^H\tilde{\boldsymbol{S}}_r)^{-1}\tilde{\boldsymbol{S}}_r^H \\ &= \boldsymbol{I} - \boldsymbol{R}^{\frac{1}{2}}\boldsymbol{S}_r(\boldsymbol{S}_r^H\boldsymbol{R}\boldsymbol{S}_r)^{-1}\boldsymbol{S}_r^H\boldsymbol{R}^{\frac{1}{2}} \end{aligned} \tag{2.52}$$

which is the orthogonal complement projection matrix of the transformed Krylov subspace $\boldsymbol{R}^{\frac{1}{2}}\mathcal{K}(\boldsymbol{R}, \boldsymbol{s}, r+1)$. Substituting (2.52) into (2.51), we have

$$\boldsymbol{w}_{\text{MF}} - \boldsymbol{w}_{r+1} = \left(\boldsymbol{R}^{-1} - \boldsymbol{S}_r(\boldsymbol{S}_r^H\boldsymbol{R}\boldsymbol{S}_r)^{-1}\boldsymbol{S}_r^H\right)\boldsymbol{s}. \tag{2.53}$$

Since $\boldsymbol{w}_{\text{MF}} = \boldsymbol{R}^{-1}\boldsymbol{s}$, and the vector $\boldsymbol{S}_r(\boldsymbol{S}_r^H\boldsymbol{R}\boldsymbol{S}_r)^{-1}\boldsymbol{S}_r^H\boldsymbol{s} \in \mathcal{K}(\boldsymbol{R}, \boldsymbol{s}, r+1)$, so

$$\boldsymbol{w}_r = \boldsymbol{S}_r(\boldsymbol{S}_r^H\boldsymbol{R}\boldsymbol{S}_r)^{-1}\boldsymbol{S}_r^H\boldsymbol{s}, \tag{2.54}$$

and

$$\boldsymbol{w}_{\text{MF}} - \boldsymbol{w}_{r+1} = \left(\boldsymbol{R}^{-1} - \boldsymbol{S}_r(\boldsymbol{S}_r^H\boldsymbol{R}\boldsymbol{S}_r)^{-1}\boldsymbol{S}_r^H\right)\boldsymbol{s}. \tag{2.55}$$

Expanding $\boldsymbol{R}^m = (\boldsymbol{R}_0 + \boldsymbol{\Delta})^m$, we have

$$\begin{aligned}\boldsymbol{R}^m &= \boldsymbol{R}_0^m + \sum_{k=1}^{m} \boldsymbol{R}_0^{m-k}\boldsymbol{\Delta}\boldsymbol{R}_0^{k-1} + o(\|\boldsymbol{\Delta}\|) \\ &\approx \boldsymbol{R}_0^m + \sum_{k=1}^{m} \boldsymbol{R}_0^{m-k}\boldsymbol{\Delta}\boldsymbol{R}_0^{k-1}. \end{aligned} \tag{2.56}$$

If the columns of $\boldsymbol{T}_r$ span the Krylov subspace $\mathcal{K}(\boldsymbol{R}_0, \boldsymbol{s}, r+1)$, then $\boldsymbol{S}_r$ can be approximated by

$$\boldsymbol{S}_r \approx \boldsymbol{T}_r + \boldsymbol{\Phi}_{\Delta} \tag{2.57}$$

where $\boldsymbol{\Phi}_\Delta$ is defined by (2.31). After substituting (2.57) into (2.55), while using a first-order expansion on $(\boldsymbol{S}_r^H \boldsymbol{R} \boldsymbol{S}_r)^{-1}$, $(\boldsymbol{S}_r^H \boldsymbol{R} \boldsymbol{S}_r)^{-1}$ can be approximated as

$$\begin{aligned}(\boldsymbol{S}_r^H \boldsymbol{R} \boldsymbol{S}_r)^{-1} &= \big((\boldsymbol{T}_r + \boldsymbol{\Phi}_\Delta)^H (\boldsymbol{R}_0 + \boldsymbol{\Delta})(\boldsymbol{T}_r + \boldsymbol{\Phi}_\Delta)\big)^{-1} \\ &\approx (\boldsymbol{T}_r^H \boldsymbol{R}_0 \boldsymbol{T}_r)^{-1} - (\boldsymbol{T}_r^H \boldsymbol{R}_0 \boldsymbol{T}_r)^{-1} \boldsymbol{\Psi}_\Delta (\boldsymbol{T}_r^H \boldsymbol{R}_0 \boldsymbol{T}_r)^{-1} \end{aligned} \tag{2.58}$$

where $\boldsymbol{\Psi}_\Delta = \boldsymbol{T}_r^H \boldsymbol{R}_0 \boldsymbol{\Phi}_\Delta + \boldsymbol{T}_r^H \boldsymbol{\Delta} \boldsymbol{T}_r + \boldsymbol{\Phi}_\Delta^H \boldsymbol{R}_0 \boldsymbol{T}_r$. Similarly using expansion on $\boldsymbol{R}^{-1}$, we have

$$\boldsymbol{R}^{-1} = (\boldsymbol{R}_0 + \boldsymbol{\Delta})^{-1} \approx \boldsymbol{R}_0^{-1} - \boldsymbol{R}_0^{-1} \boldsymbol{\Delta} \boldsymbol{R}_0^{-1}. \tag{2.59}$$

Substituting (2.58) and (2.59) into (2.55), and discarding the second and higher-order terms, we have

$$\begin{aligned}\boldsymbol{w}_{\text{MF}} - \boldsymbol{w}_{r+1} \approx &\big(\boldsymbol{R}_0^{-1} - \boldsymbol{T}_r (\boldsymbol{T}_r^H \boldsymbol{R}_0 \boldsymbol{T}_r)^{-1} \boldsymbol{T}_r^H \\ &+ \boldsymbol{T}_r (\boldsymbol{T}_r^H \boldsymbol{R} \boldsymbol{T}_r)^{-1} \boldsymbol{\Psi}_\Delta (\boldsymbol{T}_r^H \boldsymbol{R}_0 \boldsymbol{T}_r)^{-1} \boldsymbol{T}_r^H \\ &- \boldsymbol{\Phi}_\Delta (\boldsymbol{T}_r^H \boldsymbol{R}_0 \boldsymbol{T}_r)^{-1} \boldsymbol{T}_r^H - \boldsymbol{R}_0^{-1} \boldsymbol{\Delta} \boldsymbol{R}_0^{-1} \\ &- \boldsymbol{T}_r (\boldsymbol{T}_r^H \boldsymbol{R}_0 \boldsymbol{T}_r)^{-1} \boldsymbol{\Phi}_\Delta^H \big) \boldsymbol{s}. \end{aligned} \tag{2.60}$$

Since $\boldsymbol{R}_0$ is a rank-r correction matrix to $\sigma_{\text{n}}^2 \boldsymbol{I}$, the solution $\boldsymbol{R}_0^{-1} \boldsymbol{s}$ lies in the Krylov subspace $\mathcal{K}(\boldsymbol{R}_0, \boldsymbol{s}, r+1)$ [10]. Hence, $\boldsymbol{R}_0^{-\frac{1}{2}} \boldsymbol{s}$ lies in the linearly transformed Krylov subspace $\boldsymbol{R}_0^{\frac{1}{2}} \mathcal{K}(\boldsymbol{R}_0, \boldsymbol{s}, r+1)$ or the column space of $\tilde{\boldsymbol{T}}_r = \boldsymbol{R}_0^{\frac{1}{2}} \boldsymbol{T}_r$, and as such $\boldsymbol{P}_{\tilde{\boldsymbol{T}}_r}^{\perp} \boldsymbol{R}_0^{-\frac{1}{2}} \boldsymbol{s}$ is equal to zero vector, i.e.

$$\begin{aligned}&\boldsymbol{P}_{\tilde{\boldsymbol{T}}_r}^{\perp} \boldsymbol{R}_0^{-\frac{1}{2}} \boldsymbol{s} \\ &= \Big(\boldsymbol{I} - \boldsymbol{R}_0^{\frac{1}{2}} \boldsymbol{T}_r \big((\boldsymbol{R}_0^{\frac{1}{2}} \boldsymbol{T}_r)^H (\boldsymbol{R}_0^{\frac{1}{2}} \boldsymbol{T}_r)\big)^{-1} \boldsymbol{T}_r^H \boldsymbol{R}_0^{\frac{1}{2}} \Big) \boldsymbol{R}_0^{-\frac{1}{2}} \boldsymbol{s} \\ &= (\boldsymbol{R}_0^{-\frac{1}{2}} - \boldsymbol{R}_0^{\frac{1}{2}} \boldsymbol{T}_r (\boldsymbol{T}_r^H \boldsymbol{R}_0 \boldsymbol{T}_r)^{-1} \boldsymbol{T}_r^H) \boldsymbol{s} = \boldsymbol{0}. \end{aligned} \tag{2.61}$$

Since $\boldsymbol{R}_0$ is a positive-definite Hermitian matrix, left multiplying both sides of (2.61) by $\boldsymbol{R}_0^{-\frac{1}{2}}$ yields

$$\big(\boldsymbol{R}_0^{-1} - \boldsymbol{T}_r (\boldsymbol{T}_r^H \boldsymbol{R}_0 \boldsymbol{T}_r)^{-1} \boldsymbol{T}_r^H\big)\, \boldsymbol{s} = \boldsymbol{0}. \tag{2.62}$$

Substituting (2.62) into (2.60), we obtain the difference

$$\begin{aligned}&\boldsymbol{w}_{\text{MF}} - \boldsymbol{w}_{r+1} \approx \\ &\Big(\boldsymbol{T}_r (\boldsymbol{T}_r^H \boldsymbol{R}_0 \boldsymbol{T}_r)^{-1} \big(\boldsymbol{T}_r^H \boldsymbol{R}_0 \boldsymbol{\Phi}_\Delta + \boldsymbol{T}_r^H \boldsymbol{\Delta} \boldsymbol{T}_r + \boldsymbol{\Phi}_\Delta^H \boldsymbol{R}_0 \boldsymbol{T}_r\big) \\ &\times (\boldsymbol{T}_r^H \boldsymbol{R}_0 \boldsymbol{T}_r)^{-1} \boldsymbol{T}_r^H - \boldsymbol{\Phi}_\Delta (\boldsymbol{T}_r^H \boldsymbol{R}_0 \boldsymbol{T}_r)^{-1} \boldsymbol{T}_r^H \\ &- \boldsymbol{R}_0^{-1} \boldsymbol{\Delta} \boldsymbol{R}_0^{-1} - \boldsymbol{T}_r (\boldsymbol{T}_r^H \boldsymbol{R}_0 \boldsymbol{T}_r)^{-1} \boldsymbol{\Phi}_\Delta^H \Big) \boldsymbol{s}. \end{aligned} \tag{2.63}$$

Note that

$$\begin{aligned}&\boldsymbol{T}_r(\boldsymbol{T}_r^H\boldsymbol{R}_0\boldsymbol{T}_r)^{-1}\boldsymbol{T}_r^H\boldsymbol{\Delta}\boldsymbol{T}_r(\boldsymbol{T}_r^H\boldsymbol{R}_0\boldsymbol{T}_r)^{-1}\boldsymbol{T}_r^H\boldsymbol{s}\\&=\boldsymbol{R}_0^{-\frac{1}{2}}\boldsymbol{P}_{\tilde{\boldsymbol{T}}_r}\boldsymbol{R}_0^{-\frac{1}{2}}\boldsymbol{\Delta}\boldsymbol{R}_0^{-\frac{1}{2}}\boldsymbol{P}_{\tilde{\boldsymbol{T}}_r}\boldsymbol{R}_0^{-\frac{1}{2}}\boldsymbol{s}\end{aligned}\tag{2.64}$$

where $\boldsymbol{P}_{\tilde{\boldsymbol{T}}_r}=\tilde{\boldsymbol{T}}_r(\tilde{\boldsymbol{T}}_r^H\tilde{\boldsymbol{T}}_r)^{-1}\tilde{\boldsymbol{T}}_r^H$ is the orthogonal projection matrix onto the column space of $\tilde{\boldsymbol{T}}_r$. From the previous analysis, the vector $\boldsymbol{R}_0^{-\frac{1}{2}}\boldsymbol{s}$ lies in the column space of $\tilde{\boldsymbol{T}}_r$. Therefore,

$$\boldsymbol{P}_{\tilde{\boldsymbol{T}}_r}\boldsymbol{R}_0^{-\frac{1}{2}}\boldsymbol{s}=\boldsymbol{R}_0^{-\frac{1}{2}}\boldsymbol{s}.\tag{2.65}$$

Substituting (2.65) into (2.64), we have

$$\begin{aligned}&\boldsymbol{T}_r(\boldsymbol{T}_r^H\boldsymbol{R}_0\boldsymbol{T}_r)^{-1}\boldsymbol{T}_r^H\boldsymbol{\Delta}\boldsymbol{T}_r(\boldsymbol{T}_r^H\boldsymbol{R}_0\boldsymbol{T}_r)^{-1}\boldsymbol{T}_r^H\boldsymbol{s}\\&=\boldsymbol{R}_0^{-\frac{1}{2}}\boldsymbol{P}_{\tilde{\boldsymbol{T}}_r}\boldsymbol{R}_0^{-\frac{1}{2}}\boldsymbol{\Delta}\boldsymbol{R}_0^{-1}\boldsymbol{s}\end{aligned}\tag{2.66}$$

and

$$\begin{aligned}&\big(\boldsymbol{R}_0^{-1}\boldsymbol{\Delta}\boldsymbol{R}_0^{-1}-\boldsymbol{T}_r(\boldsymbol{T}_r^H\boldsymbol{R}_0\boldsymbol{T}_r)^{-1}\boldsymbol{T}_r^H\boldsymbol{\Delta}\boldsymbol{T}_r\\&\times(\boldsymbol{T}_r^H\boldsymbol{R}_0\boldsymbol{T}_r)^{-1}\boldsymbol{T}_r^H\big)\boldsymbol{s}\\&=\boldsymbol{R}_0^{-\frac{1}{2}}\boldsymbol{P}_{\tilde{\boldsymbol{T}}_r}^{\perp}\boldsymbol{R}_0^{-\frac{1}{2}}\boldsymbol{\Delta}\boldsymbol{R}_0^{-1}\boldsymbol{s}.\end{aligned}\tag{2.67}$$

Similarly, it can be be proved that

$$\begin{aligned}&\Big(\boldsymbol{\Phi}_{\Delta}(\boldsymbol{T}_r^H\boldsymbol{R}_0\boldsymbol{T}_r)^{-1}\boldsymbol{T}_r^H\\&-\boldsymbol{T}_r(\boldsymbol{T}_r^H\boldsymbol{R}_0\boldsymbol{T}_r)^{-1}\boldsymbol{T}_r^H\boldsymbol{R}_0\boldsymbol{\Phi}_{\Delta}(\boldsymbol{T}_r^H\boldsymbol{R}_0\boldsymbol{T}_r)^{-1}\boldsymbol{T}_r^H\Big)\boldsymbol{s}\\&=\boldsymbol{R}_0^{-\frac{1}{2}}\boldsymbol{P}_{\tilde{\boldsymbol{T}}_r}^{\perp}\boldsymbol{R}_0^{\frac{1}{2}}\boldsymbol{\Phi}_{\Delta}(\boldsymbol{T}_r^H\boldsymbol{R}_0\boldsymbol{T}_r)^{-1}\boldsymbol{T}_r^H\boldsymbol{s}\end{aligned}\tag{2.68}$$

and

$$\begin{aligned}&\Big(\boldsymbol{T}_r(\boldsymbol{T}_r^H\boldsymbol{R}_0\boldsymbol{T}_r)^{-1}\boldsymbol{\Phi}_{\Delta}^H\\&-\boldsymbol{T}_r(\boldsymbol{T}_r^H\boldsymbol{R}_0\boldsymbol{T}_r)^{-1}\boldsymbol{\Phi}_{\Delta}^H\boldsymbol{R}_0\boldsymbol{T}_r(\boldsymbol{T}_r^H\boldsymbol{R}_0\boldsymbol{T}_r)^{-1}\boldsymbol{T}_r^H\Big)\boldsymbol{s}\\&=\boldsymbol{T}_r(\boldsymbol{T}_r^H\boldsymbol{R}_0\boldsymbol{T}_r)^{-1}\boldsymbol{\Phi}_{\Delta}^H\boldsymbol{P}_{\tilde{\boldsymbol{T}}_r}^{\perp}\boldsymbol{R}_0^{-\frac{1}{2}}\boldsymbol{s}=\boldsymbol{0}.\end{aligned}\tag{2.69}$$

Substituting (2.67)-(2.69) into (2.63), we have

$$\begin{aligned}\boldsymbol{w}_{\text{MF}}-\boldsymbol{w}_{r+1}&\approx\boldsymbol{R}_0^{-\frac{1}{2}}\boldsymbol{P}_{\tilde{\boldsymbol{T}}_r}^{\perp}\boldsymbol{R}_0^{-\frac{1}{2}}\\&\quad\big(\boldsymbol{\Delta}\boldsymbol{R}_0^{-1}+\boldsymbol{R}_0\boldsymbol{\Phi}_{\Delta}(\boldsymbol{T}_r^H\boldsymbol{R}_0\boldsymbol{T}_r)^{-1}\boldsymbol{T}_r^H\big)\,\boldsymbol{s}\\&=\boldsymbol{R}_0^{-\frac{1}{2}}\boldsymbol{P}_{\tilde{\boldsymbol{T}}_r}^{\perp}\boldsymbol{R}_0^{-\frac{1}{2}}\boldsymbol{d}\end{aligned}\tag{2.70}$$

where $\boldsymbol{d}$ is defined by (2.30). Q.E.D.

Chapter 3

Conjugate Gradient Parametric Matched Filter

3.1 Introduction

With extra spatial information provided by multiple sensors, higher performance of signal detection can be achieved (than a single-sensor system), especially in detection of signals buried in a background of directional jammers and space-time correlated clutter. A widely explored technology for multichannel signal detection is space-time adaptive processing (STAP) [1], first proposed by Brennan, Reed and Mallett [2]. Most STAP-based methods, such as the adaptive matched filter (AMF) [3] and Kelly's generalized likelihood ratio test (GLRT) [4], need to invert a large space-time covariance matrix. These methods require not only a large number of independent, identically distributed, signal-free training data to estimate the matrix, but they also incur a high computational cost for matrix estimation and inversion.

A parametric STAP detector based on a multichannel autoregressive (AR) disturbance model has been proposed for airborne radar applications [16, 21] to reduce both the training data requirement and computation load. This method is called the parametric adaptive matched filter (PAMF) [16]. While the PAMF detector has been found to yield exceptional performance with significantly reduced training and computational requirements when compared with fully adaptive STAP detectors, the connections between the PAMF and other reduced-dimensional or partially adaptive STAP detectors [1], which have similar benefits in training and complexity, remains unclear.

This chapter aims to provide some insights into this problem by employing the conjugate-gradient (CG) method to solve the linear prediction problem underlying the temporal whitening phase of the PAMF detector. Our choice of the CG method is motivated by several factors. First, as will be shown, the CG algorithm naturally leads to a subspace interpretation of the PAMF detector, and offers a connection to the other reduced-rank STAP detectors. Second, the CG method is a computational efficient algorithm to solve the linear prediction problem underlying the PAMF detector. In particular, for airborne radar applications, due to an inherent structure of

the disturbance covariance matrix, the CG algorithm can usually achieve convergence using only a few iterations, thus providing significant computational saving. Third, since the disturbance covariance matrix has a block-Toeplitz matrix structure, preconditioning methods(e.g., [10, 22, 23]) can be employed, which are very effective in further speeding up the convergence rate in CG iterations, while adding up only a modest computational overhead per iteration (due to the block-Toeplitz structure). Finally, as a by-product, we show that the CG algorithm also yields a new and computationally efficient AR model order selection method that can be integrated with the CG iterations.

The remainder of this chapter is organized as follows. The signal detection problem is introduced in Section 3.2. A brief review of the matched filter (MF) and parametric matched filter (PMF) detectors is provided in Section 3.3. In Section 3.4, the CG versions of MF (CG-MF) and PMF (CG-PMF) and a CG-based model order selection method are proposed. The convergence rate of CG in airborne radar applications, along with a preconditioned CG-PMF (PCG-PMF) detector, are also discussed there. In Section 3.5, we consider the adaptive case and present a new model order-selection CG-PAMF (OSCG-PAMF) detector, when both the AR model order and coefficients are unknown. The performance of the proposed class of CG-PMF and CG-PAMF detectors is illustrated by numerical results in Section 3.6. Finally conclusions are summarized in Section 3.7.

Vectors and matrices are denoted by boldface lower-case and upper-case letters, respectively. Transpose, complex conjugate and complex conjugate transpose are respectively represented by $(\cdot)^T$, $(\cdot)^*$ and $(\cdot)^H$. $\mathbb{C}$ and $\mathbb{R}$ denote the complex and real number fields. $\mathcal{CN}(\boldsymbol{\mu}, \mathbf{R})$ denotes an additive multivariate Gaussian random variable with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{R}$.

3.2 Data Model

Consider a received J-channel sequence $\{\mathbf{x}(n)|n = 1, 2, \cdots, N\}$ corrupted by a space-time correlated disturbance random process $\mathbf{c}(n)$. The detection problem involves the following binary hypotheses:

$$\begin{aligned} H_0 &: \mathbf{x}(n) = \mathbf{c}(n) \\ H_1 &: \mathbf{x}(n) = a\mathbf{s}(n) + \mathbf{c}(n) \end{aligned} \tag{3.1}$$

where $\mathbf{s}(n)$ is a known J-channel signal and a is its deterministic but unknown complex amplitude. All vectors in (3.1) are $J \times 1$ vectors. For convenience of later discussions, define the following vectors in descending order: $\mathbf{s} = [\mathbf{s}^T(N), \mathbf{s}^T(N-1), \cdots, \mathbf{s}^T(1)]^T$, $\mathbf{c} = [\mathbf{c}^T(N), \mathbf{c}^T(N-1), \cdots, \mathbf{c}^T(1)]^T$, $\mathbf{x} = [\mathbf{x}^T(N), \mathbf{x}^T(N-1), \cdots, \mathbf{x}^T(1)]^T$. It is standard to assume that the disturbance $\mathbf{c}$ is a Gaussian random vector with zero-mean and space-time covariance matrix $\mathbf{R}_c \in \mathbb{C}^{JN \times JN}$, while the signal vector $\mathbf{s}(n)$ is deterministic (Swerling 0 target). Based on these assumptions, $\mathbf{x} \sim \mathcal{CN}(a\mathbf{s}, \mathbf{R}_c)$, where $a = 0$ under H_0 and $a \neq 0$ under H_1.

In STAP, the signal $\mathbf{s}$ is known as the space-time steering vector:

$$\mathbf{s} = \mathbf{s}_t \otimes \mathbf{s}_s \tag{3.2}$$

where $\mathbf{s}_t$ and $\mathbf{s}_s$ denote the temporal steering vector and spatial steering vector, respectively, and $\otimes$ denotes the Kronecker product. For a side-looking uniform linear array (ULA), we have $\mathbf{s}_t = (1/\sqrt{N})[e^{i2\pi(N-1)f_d}, \cdots, e^{i2\pi f_d}, 1]^T$ with a normalized Doppler frequency f_d and $\mathbf{s}_s = (1/\sqrt{J})[e^{i2\pi(J-1)f_s}, \cdots, e^{i2\pi f_s}, 1]^T$ with a normalized spatial frequency f_s. Practically, the true disturbance covariance matrix $\mathbf{R}_c$ is unknown, and often an estimate can be obtained from the secondary data:

$$\hat{\mathbf{R}}_c = \frac{1}{K}\sum_{k=1}^{K} \mathbf{c}_k \mathbf{c}_k^H \tag{3.3}$$

where $\mathbf{c}_k, k = 1, 2 \cdots, K$, denote the secondary data vectors assumed to be signal-free. According to the well-known "RMB" rule [24], we need $K \geq 2JN - 3$ so that the average output signal-to-interference-plus-noise ratio (SINR) loss caused by covariance estimation error is less than 3 dB. Detectors with an estimated covariance matrix are often called adaptive methods.

3.3 MF and PMF

Assuming a known $\mathbf{R}_c$, the matched filter (MF) is obtained by maximizing the output SINR of a linear receiver or the generalized likelihood ratio (GLR). The test is given by (e.g., [3]):

$$\frac{|\mathbf{s}^H \mathbf{R}_c^{-1}\mathbf{x}|^2}{\mathbf{s}^H \mathbf{R}_c^{-1}\mathbf{s}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{\text{MF}} \tag{3.4}$$

where η_{MF} is the threshold of the MF. Equation (3.4) is the well known matched subspace detector for a rank-1 signal in colored noise. Consequently, it offers unbeatable performance for the detection problem considered in equation (3.1).

For ease of exposition, the MF can also be represented by using a structure of temporal whitening cascaded with spatial whitening arising from a block LDU decomposition of the disturbance covariance matrix [16]. This form of MF is given by

$$\frac{|(\mathbf{Q}^{-1/2}\mathbf{L}^{-1}\mathbf{s})^H(\mathbf{Q}^{-1/2}\boldsymbol{\varepsilon})|^2}{(\mathbf{Q}^{-1/2}\mathbf{L}^{-1}\mathbf{s})^H(\mathbf{Q}^{-1/2}\mathbf{L}^{-1}\mathbf{s})} = \frac{|\tilde{\mathbf{s}}^H\boldsymbol{\nu}|^2}{\tilde{\mathbf{s}}^H\tilde{\mathbf{s}}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{\text{MF}} \tag{3.5}$$

where $\mathbf{Q} \in \mathbb{C}^{JN\times JN}$ is a block-diagonal matrix with Hermitian matrices $\mathbf{Q}(n), n = 1, 2, \cdots, N$, along the main block diagonal, and $\mathbf{L} \in \mathbb{C}^{JN\times JN}$ is a lower block-triangular matrix with $J \times J$ identity matrices along the main block diagonal. Both $\mathbf{L}$ and $\mathbf{Q}$ come from a block LDU decomposition of the disturbance covariance matrix $\mathbf{R}_c = \mathbf{L}\mathbf{Q}\mathbf{L}^H$. Finally,

$$\boldsymbol{\varepsilon}(n) = \mathbf{x}(n) - \sum_{p=1}^{(n-1)} \mathbf{A}_n^H(p)\mathbf{x}(n-p) \tag{3.6}$$

$$\boldsymbol{\nu}(n) = \mathbf{Q}^{-1/2}(n)\boldsymbol{\varepsilon}(n) \tag{3.7}$$

$$\tilde{\mathbf{s}}(n) = \mathbf{Q}^{-1/2}(n)\left[\mathbf{s}(n) - \sum_{p=1}^{(n-1)} \mathbf{A}_n^H(p)\mathbf{s}(n-p)\right] \tag{3.8}$$

where $\mathbf{A}_n^H(p) \in \mathbb{C}^{J\times J}$ is a block element of $\mathbf{L}^{-1}$ located at the $(n-p)$-th block column and the n-th block row. Due to the fact that there is no performance penalty for the prewhitening of the interference [25, Ch.6], (3.5) is equivalent to (3.4).

If the disturbance $\mathbf{c}(n)$ is stationary in time, the MF can be simplified. A parametric matched filter (PMF) was introduced in [16] by modeling the disturbance as a stationary P-th-order multichannel autoregressive (AR) process. Specifically,

$$\mathbf{c}(n) = \sum_{p=1}^{P} \mathbf{A}^H(p)\mathbf{c}(n-p) + \boldsymbol{\varepsilon}_P(n) \tag{3.9}$$

where $\mathbf{A}^H(p), p = 1, 2, \cdots, P$, is the p-th AR matrix coefficient of linear prediction, and $\boldsymbol{\varepsilon}_P(n)$ is the temporally white noise with a spatial covariance matrix $\mathbf{Q}_P$. The PMF test is given by [16]

$$\frac{|\sum_{n=P+1}^{N} \tilde{\mathbf{s}}_P^H(n)\boldsymbol{\nu}_P(n)|^2}{\sum_{n=P+1}^{N} \tilde{\mathbf{s}}_P^H(n)\tilde{\mathbf{s}}_P(n)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{\text{PMF}} \tag{3.10}$$

where $\boldsymbol{\nu}_P(n) = \mathbf{Q}_P^{-1/2}\boldsymbol{\varepsilon}_P(n)$ and

$$\tilde{\mathbf{s}}_P(n) = \mathbf{Q}_P^{-1/2}\left[\mathbf{s}(n) - \sum_{p=1}^{P} \mathbf{A}^H(p)\mathbf{s}(n-p)\right] \tag{3.11}$$

for $n = P+1, \cdots, N$. In practice, the model order P and the AR coefficients $\{\mathbf{A}(p)\}$ are unknown and hence estimated from the secondary data and/or primary data. Different estimators lead to different versions and implementations of the PAMF detector [16, 17].

3.4 CG-MF and CG-PMF

In this section, we discuss alternative implementations of the MF and PMF via the CG algorithm. The resulting detectors are referred to as the CG-MF and CG-PMF detectors, respectively, for brevity. We start from the CG-MF, which also sets the basis for the CG-PMF. The latter, by assuming that the disturbance $\mathbf{c}(n)$ is temporally stationary, is a computationally simplified version of the CG-MF. The link between the PMF and CG as developed in the sequel reveals the PMF as a reduced-dimensional subspace detector. In this section, we assume knowledge of the covariance matrix of the disturbance signal. An adaptive versions of the CG-PMF (i.e., CG-PAMF) is discussed in Section 3.5.

3.4.1 Conjugate-Gradient MF

The MF detector, as shown in Section 3.3, can be derived from a time-varying linear prediction process. Specifically, consider the problem of linearly predicting the n-th sample $\mathbf{x}(n)$ under H_0 from all prior received samples $\mathbf{x}(n-1), \mathbf{x}(n-2), \ldots, \mathbf{x}(1)$ (cf. (3.9))

$$\mathbf{x}(n) = \mathbf{B}^H(n)\mathbf{y}(n) + \boldsymbol{\varepsilon}(n) \tag{3.12}$$

where $\mathbf{B}(n) = [\mathbf{A}_n^H(1), \mathbf{A}_n^H(2), \cdots, \mathbf{A}_n^H(n-1)]^H = [\mathbf{B}_1(n), \mathbf{B}_2(n), \cdots, \mathbf{B}_J(n)] \in \mathbb{C}^{J(n-1)\times J}$ denotes the $(n-1)$-st order time-varying multichannel linear prediction filter, and $\mathbf{y}(n) = [y_n(1), y_n(2), \cdots, y_n(J(n-1)]^T = [\mathbf{x}^T(n-1), \mathbf{x}^T(n-2), \cdots, \mathbf{x}^T(1)]^T$ contains all $n-1$ previously received data vectors. It is noted that the above time-varying linear predictor grows in its filter order or size with n. The multichannel linear predictor is equivalent to J linear predictors:

$$x_j(n) = \mathbf{B}_j^H(n)\mathbf{y}(n) + \varepsilon_j(n), \quad j = 1, 2, \cdots, J \tag{3.13}$$

where $\mathbf{B}_j(n)$ is a $J(n-1)$-dimensional vector which contains the cross-channel correlation information associated with the j-th channel. The optimum linear predictor can be obtained by solving the Wiener-Hopf equations:

$$\mathbf{R}_{\mathbf{y}}(n)\mathbf{B}_j(n) = \mathbf{R}_j(n), \quad j = 1, 2, \cdots, J \tag{3.14}$$

where $\mathbf{R}_{\mathbf{y}}(n) = E[\mathbf{y}(n)\mathbf{y}^H(n)] \in \mathbb{C}^{J(n-1)\times J(n-1)}$ and $\mathbf{R}_j(n) = E[\mathbf{y}(n)x_j^*(n)] \in \mathbb{C}^{J(n-1)\times 1}$. Again, note that the size of the Wiener-Hopf equation grows with n.

To obtain a temporally whitened sequence $\boldsymbol{\varepsilon}(n)$ for MF detection (cf. (3.6)), the above linear prediction process has to be performed multiple times, starting from $n = 2$ to $n = N$. For each n, we need to solve a Wiener-Hopf equation of the form (3.14). While there are various solvers to the linear Wiener-Hopf equation, we consider using the conjugate gradient (CG) method, which has several properties such as fast convergence, a direct link to the Krylov subspace [10], and a built in model order selection capability. Additional remarks on such aspects are provided shortly.

The recursive procedure involved for the determination of the linear predictors is described as follows (also see (3.9)).

for $n = 2$ to N **do**
  for $j = 1$ to J **do**
    Initialization. Initialize the conjugate-direction vector $\mathbf{D}_{0,j}(n)$, gradient vector $\boldsymbol{\gamma}_{1,j}(n)$ and initial solution $\mathbf{B}_{0,j}(n)$:

$$\mathbf{D}_{1,j}(n) = -\boldsymbol{\gamma}_{1,j}(n) = \mathbf{R}_j(n) \tag{3.15}$$

$$\mathbf{B}_{0,j}(n) = \mathbf{0}. \tag{3.16}$$

    for $k = 1, 2, \cdots$, till convergence ($k \le J(n-1)$) **do**
      Update the step size $\alpha_{k,j}$:

$$\alpha_{k,j}(n) = \frac{\|\boldsymbol{\gamma}_{k,j}(n)\|^2}{\mathbf{D}_{k,j}^H(n)\mathbf{R}_{\mathbf{y}}(n)\mathbf{D}_{k,j}(n)}. \tag{3.17}$$

      Update the solution $\mathbf{B}_{k,j}$:

$$\mathbf{B}_{k,j}(n) = \mathbf{B}_{k-1,j}(n) + \alpha_{k,j}(n)\mathbf{D}_{k,j}(n). \tag{3.18}$$

      Update the gradient vector $\boldsymbol{\gamma}_{k+1,j}$:

$$\boldsymbol{\gamma}_{k+1,j}(n) = \boldsymbol{\gamma}_{k,j}(n) + \alpha_{k,j}(n)\mathbf{R}_{\mathbf{y}}(n)\mathbf{D}_{k,j}(n). \tag{3.19}$$

Figure 3.1: Time-varying linear prediction in the conjugate-gradient MF detector.

Update the conjugate-direction vector $\mathbf{D}_{k+1,j}$:

$$\mathbf{D}_{k+1,j}(n) = \mathbf{D}_{k,j}(n)\frac{\|\boldsymbol{\gamma}_{k+1,j}(n)\|^2}{\|\boldsymbol{\gamma}_{k,j}(n)\|^2} - \boldsymbol{\gamma}_{k+1,j}(n). \tag{3.20}$$

end for
end for
end for

Let $\mathbf{B}(n)$ be the multichannel linear predictor formed from $\mathbf{B}_{k,j}$ after convergence. Then, $\mathbf{B}(n)$ can be used to whiten $\mathbf{x}(n)$ to produce a temporally whitened sequence $\boldsymbol{\varepsilon}(n)$. The spatial covariance matrix $\mathbf{Q}(n)$ of $\boldsymbol{\varepsilon}(n)$ is given by (cf. (3.15))

$$\begin{aligned}\mathbf{Q}(n) &= E[\boldsymbol{\varepsilon}(n)\boldsymbol{\varepsilon}^H(n)] \\ &= \mathbf{R}_{\mathbf{x}}(n) - \mathbf{B}^H(n)\mathbf{R}_{\mathbf{yx}}(n)\end{aligned} \tag{3.21}$$

where $\mathbf{R}_{\mathbf{x}}(n) = E[\mathbf{x}(n)\mathbf{x}^H(n)] \in \mathbb{C}^{J\times J}$, and $\mathbf{R}_{\mathbf{yx}}(n) = E[\mathbf{y}(n)\mathbf{x}^H(n)] \in \mathbb{C}^{J(n-1)\times J}$, which is used for further spatial whitening [16].

Fig. 3.1 depicts the CG-MF detector that produces the n-th sample of the temporally whitened sequence $\varepsilon_j(n)$ for the j-th channel, where $\mathbf{D}_{k,j}(n) = [\mathbf{D}_{1,j}(n), \mathbf{D}_{2,j}(n), \cdots, \mathbf{D}_{k,j}(n)]$ is the conjugate-direction matrix. CG iterations lead to a set of linearly independent vectors $\mathbf{D}_{1,j}(n), \ldots, \mathbf{D}_{k,j}(n)$ that are *conjugate orthogonal*, i.e.

$$\mathbf{D}_{k,j}^H(n)\mathbf{R}_{\mathbf{y}}(n)\mathbf{D}_{l,j}(n) = 0, \quad k \neq l. \tag{3.22}$$

The output of the k-th iteration is given by

$$\mathbf{B}_{k,j}(n) = \sum_{m=1}^{k} \alpha_{m,j}(n)\mathbf{D}_{m,j}(n) \tag{3.23}$$

which is a vector in the k-dimensional vector space spanned by the conjugate-direction vectors $\{\mathbf{D}_{m,j}(n), m = 1, 2, \cdots, k\}$. The iterative procedure for the prediction of the n-th sample $x_j(n)$, which involves a $J(n-1)$-st order linear predictor, converges after at most $J(n-1)$ iterations. The final solution $\mathbf{B}_j(n)$ lies in a $J(n-1)$-dimensional vector space.

Figure 3.2: Time-invariant linear prediction in the conjugate-gradient PMF detector.

3.4.2 Conjugate-Gradient PMF with Known AR Model Order

If the disturbance signal can be approximated as a temporally wide-sense stationary (WSS) multichannel AR process, the linear prediction problem of the previous subsection can be significantly simplified. Specifically, suppose the disturbance is an AR(P) process with model order P. In this case, the optimum linear predictor for the n-th sample $\mathbf{x}(n)$ requires only P most recently received samples (as opposed to all past samples) and the prediction filter is time-invariant with a fixed size (as opposed to time-varying with a growing size) [26]:

$$\mathbf{x}(n) = \mathbf{B}^H \mathbf{y}_P(n) + \boldsymbol{\varepsilon}_P(n) \tag{3.24}$$

where the fixed P-th order linear predictor

$$\mathbf{B} = [\mathbf{A}^H(1), \mathbf{A}^H(2), \cdots, \mathbf{A}^H(P)]^H = [\mathbf{B}_1, \mathbf{B}_2, \cdots, \mathbf{B}_J] \in \mathbb{C}^{JP \times J} \tag{3.25}$$

is composed of the AR coefficient matrices $\{\mathbf{A}^H(p)\}$ (cf. (3.9)),

$$\mathbf{y}_P(n) = [y_n(1), y_n(2), \cdots, y_n(JP)]^T = [\mathbf{x}^T(n-1), \mathbf{x}^T(n-2), \cdots, \mathbf{x}^T(n-P)]^T \tag{3.26}$$

denotes the regression data vector, and $n > P$. Again, it is convenient to express the above multichannel linear predictor as J scalar linear predictors:

$$x_j(n) = \mathbf{B}_j^H \mathbf{y}_P(n) + \varepsilon_{P,j}(n), \quad j = 1, 2, \cdots, J. \tag{3.27}$$

The structure of temporal whitening via linear prediction for the PMF detector is shown in Fig. 3.2.

The solution to the scalar linear prediction problem can be obtained by solving the following Wiener-Hopf equation

$$\mathbf{R}_{\mathbf{y}} \mathbf{B}_j = \mathbf{R}_j, \quad j = 1, 2, \cdots, J \tag{3.28}$$

where $\mathbf{R}_{\mathbf{y}} = E[\mathbf{y}_P(n)\mathbf{y}_P^H(n)] \in \mathbb{C}^{JP \times JP}$ and $\mathbf{R}_j = E[\mathbf{y}_P(n)x_j^*(n)] \in \mathbb{C}^{JP \times 1}$. It should be noted that unlike the MF detector, the above Wiener-Hopf is time-invariant, has a fixed size, and needs to be solved only once. The resulting solution $\mathbf{B}_j$ can be used to whiten the entire received signal $\mathbf{x}(n)$ for $n > P$. The CG algorithm can also be applied to solve (3.28), and the resulting detector is referred to as the CG-PMF detector. Since only one fixed-sized Wiener-Hopf equation needs to be solved, the CG-PMF detector is computationally much simpler. Specifically, the outer loop

for varying n as discussed in Section 3.4.1 vanishes, and only the conjugate-gradient processing with $n = P + 1$ is needed.

Remark: The iterative procedure of CG converges after at most JP iterations for the CG-PMF. As a result, the final solution $\mathbf{B}_j$ lies in a JP-dimensional vector space spanned by the conjugate direction vectors $\mathbf{D}_{k,j}, k = 1, 2, \ldots, JP$, or equivalently, the JP-dimensional Krylov subspace [10]:

$$\mathcal{K}(\mathbf{R}_j, \mathbf{R}_\mathbf{y}, JP) = \text{span}\left\{\mathbf{R}_j, \mathbf{R}_\mathbf{y}\mathbf{R}_j, \cdots, \mathbf{R}_\mathbf{y}^{JP-1}\mathbf{R}_j\right\}. \tag{3.29}$$

This shows that the PMF is a reduced JP-dimensional solution, as opposed to the full JN-dimensional MF detector. The same conclusion applies to the adaptive version CG-PAMF detector discussed in Section 3.5.

3.4.3 Model Order Selection by CG

In practice, the AR model order P of the disturbance is often unknown and has to be estimated. A practical approach is to choose an upper bound $\bar{P}$ for P, and use the CG algorithm to solve the following Wiener-Hopf equation

$$\mathbf{R}_\mathbf{y}^{(\bar{P})}\mathbf{B}_j^{(\bar{P})} = \mathbf{R}_j^{(\bar{P})}, \quad j = 1, 2, \cdots, J \tag{3.30}$$

where $\mathbf{R}_\mathbf{y}^{(\bar{P})} = E[\mathbf{y}_{\bar{P}}(n)\mathbf{y}_{\bar{P}}^H(n)] \in \mathbb{C}^{J\bar{P}\times J\bar{P}}$ and $\mathbf{R}_j^{(\bar{P})} = E[\mathbf{y}_{\bar{P}}(n)x_j^*(n)] \in \mathbb{C}^{J\bar{P}\times 1}$. The CG iterative procedure will converge after at most $J\bar{P}$ iterations with $\mathbf{B}_j^{(\bar{P})} = [\mathbf{B}_j^T, \mathbf{0}_{J(\bar{P}-P)\times 1}^T]^T$. However, with a loosely determined upper bound $\bar{P}$, it is often necessary for the sake of reducing computational complexity to stop the CG iterations before it reaches the maximum number of iterations. In this section, we propose a model order selection method for use with the CG algorithm, which is based on the following result.

Lemma 2 *Suppose the disturbance in* (3.1) *is a J-channel AR(P) process. Let $\mathbf{B}_{k,j}^{(\bar{P})} \in \mathbb{C}^{J\bar{P}\times 1}$ be the solution to* (3.30) *obtained by CG at the k-th iteration, where $k = Jp$ and $p \le \bar{P}$. Let $\mathbf{B}_j^{(p)} \in \mathbb{C}^{Jp\times 1}$ be the solution to $\mathbf{R}_\mathbf{y}^{(p)}\mathbf{B}_j^{(p)} = \mathbf{R}_j^{(p)}$. Then we have*

$$\mathbf{B}_{k,j}^{(\bar{P})} = \mathbf{W}_{k,j}\mathbf{B}_j^{(p)}, \text{ when } p = P \tag{3.31}$$

where $\mathbf{W}_{k,j} = \mathbf{D}_{k,j}^{(\bar{P})}\bar{\mathbf{D}}_{k,j}^H$, $\mathbf{D}_{k,j}^{(\bar{P})} = [\mathbf{D}_{1,j}^{(\bar{P})}, \mathbf{D}_{2,j}^{(\bar{P})}, \cdots, \mathbf{D}_{k,j}^{(\bar{P})}]$ is the conjugate-direction matrix, and $\bar{\mathbf{D}}_{k,j}$ is a $k \times k$ matrix composed of the first k rows of $\tilde{\mathbf{D}}_{k,j} = [\tilde{\mathbf{D}}_{1,j}, \tilde{\mathbf{D}}_{2,j}, \cdots, \tilde{\mathbf{D}}_{k,j}]$ with

$$\tilde{\mathbf{D}}_{k,j} = \frac{\mathbf{R}_\mathbf{y}^{(\bar{P})}\mathbf{D}_{k,j}^{(\bar{P})}}{\mathbf{D}_{k,j}^{(\bar{P})\,H}\mathbf{R}_\mathbf{y}^{(\bar{P})}\mathbf{D}_{k,j}^{(\bar{P})}} \tag{3.32}$$

proof See Appendix 3.8. Q.E.D.

From (3.31), when $p = P$, the $J\bar{P} \times JP$ matrix $\mathbf{W}_{JP,j}$ transforms $\mathbf{B}_j$ in $\mathcal{K}(\mathbf{R}_j, \mathbf{R}_\mathbf{y}, JP)$, which is generated by CG-PMF with a known P, to $\mathbf{B}_{JP,j}^{(\bar{P})}$ in $\mathcal{K}(\mathbf{R}_j, \mathbf{R}_\mathbf{y}^{(\bar{P})}, JP)$, which is generated by CG-PMF with an unknown P. So the PMF AR coefficient vector $\mathbf{B}_j$ is completely determined by the truncated solution $\mathbf{B}_{JP,j}^{(\bar{P})}$ of CG with an unknown P after JP iterations.

We now explain how to use Lemma 2 for AR model order selection in CG-PMF. Define the residue vector

$$\boldsymbol{\epsilon}_{k,j} = \mathbf{B}_{k,j}^{(\bar{P})} - \mathbf{D}_{k,j}^{(\bar{P})} \bar{\mathbf{D}}_{k,j}^H \mathbf{B}_j^{(p)}, \quad \text{for } k = Jp. \tag{3.33}$$

According to (3.31), $\boldsymbol{\epsilon}_{k,j} = \mathbf{0}$ when $k = JP$, so the norm of $\boldsymbol{\epsilon}_{k,j}$ can be used for model order selection. However, since the Wiener solution $\mathbf{B}_j^{(p)}$ is practically unknown, $\boldsymbol{\epsilon}_{k,j}$ cannot be directly computed from (3.33). We propose an approach to replace $\mathbf{B}_j^{(p)}$ in (3.33) by the truncated solution composed of the first k elements of $\mathbf{B}_{k,j}^{(\bar{P})}$, which can be considered as an approximation of $[\mathbf{B}_j^T, \mathbf{0}_{J(\bar{P}-P)\times 1}^T]^T$

$$\hat{\boldsymbol{\epsilon}}_{k,j} = \mathbf{B}_{k,j}^{(\bar{P})} - \mathbf{D}_{k,j}^{(\bar{P})} \bar{\mathbf{D}}_{k,j}^H \bar{\mathbf{B}}_{k,j}^{(\bar{P})} \tag{3.34}$$

where $\bar{\mathbf{B}}_{k,j}^{(\bar{P})}$ contains the first $k = Jp$ elements of $\mathbf{B}_{k,j}^{(\bar{P})}$. Our CG-based model order selection procedure is summarized as follows.

- ***Step 1:*** Select an upper bound $\bar{P}$ for the model order. One such an upper bound suggested in [16] for STAP detection is

$$\bar{P} = \max \left\{ \left\lfloor \frac{3\sqrt{N}}{J} \right\rfloor \right\} \tag{3.35}$$

  where $\lfloor \cdot \rfloor$ rounds a real-valued number towards zero.

- ***Step 2:*** Use the CG algorithm to solve the Wiener-Hopf equation (3.30).

  - ***Step 2.1:*** Following every J iterations of the CG algorithm, compute the average residue over J channels:

$$\bar{\hat{\epsilon}}_k^2 = \frac{1}{J} \sum_{j=1}^{J} \left\| \hat{\boldsymbol{\epsilon}}_{k,j} \right\|^2, \quad k = J, 2J, \ldots \tag{3.36}$$

  - ***Step 2.2:*** If $\bar{\hat{\epsilon}}_{Jp}^2$ is smaller than a specified tolerance level, then stop the CG iteration, and the estimated AR model order is $\hat{P} = p$.

The advantage of the above CG-based model order selection method is that it does not require full iterations of the CG algorithm and is efficient. The complexity of the CG algorithm with full iterations is in the same order as that of computing the inverse of $\mathbf{R}_\mathbf{y}^{(\bar{P})}$, which is $\mathcal{O}(J^3\bar{P}^3)$, while the complexity of using the CG-based order selection method, is $\mathcal{O}(J^3 P \bar{P}^2)$. This is because the latter only requires JP iterations to determine the model order, and the additional complexity

in each J iterations for (3.34) is the complexity of two matrix-vector multiplications, which is $2(J\bar{P})^2$. So the total complexity is $\mathcal{O}(JP(J\bar{P})^2 + 2P(J\bar{P})^2) \approx \mathcal{O}(J^3P\bar{P}^2)$. Next, we compare the computational complexity of the CG-PMF with an unknown P with the complexity of the eigencanceler [5], which is a standard eigen-subspace detector. The eigencanceler method, has a complexity of $9J^3N^3$ by using the symmetric QR algorithm to obtain the eigen-subspace and its corresponding eigen-values [10]. Since $\bar{P} \leq N$, and generally $P \ll N$, the complexity of CG-PMF is much lower than eigencanceler.

3.4.4 Convergence in Airborne Radar Applications

One important property of the CG algorithm is its fast convergence. In general, it takes no more than JP iterations to solve the linear equation (3.28) [10]. Even faster convergence is possible if the covariance matrix of the disturbance has some specific structure. In particular, if the covariance matrix is a rank-r_{c} correction of an identity matrix:

$$\mathbf{R}_{\mathbf{y}} = \mathbf{R}_{\text{i}} + \sigma_{\text{n}}^2\mathbf{I} \tag{3.37}$$

where $\mathbf{R}_{\text{i}}$ is a rank-r_{c} positive semi-definite matrix, then the CG algorithm converges in at most $r_{\text{c}} + 1$ iterations [10].

In airborne radar applications, the disturbance covariance matrix often consists of two components, namely a low-rank $\mathbf{R}_{\text{i}}$ due to the clutter and jamming and a scaled identity $\sigma_{\text{n}}^2\mathbf{I}$ due to the white noise, where σ_{n}^2 denotes the noise variance. The rank r_{c} is typically much smaller than the joint spatio-temporal dimension JN. Specifically, if the disturbance is primarily due to ground clutter and thermal noise, then according to Brennan's rule [2], the rank of the clutter covariance matrix for the full-dimensional MF is approximately

$$r_{\text{c,full}} \approx \lceil J + (N-1)\beta \rceil \tag{3.38}$$

where $\beta = 2v_{\text{g}}T_{\text{r}}/d$, v_{g} is the platform velocity, T_{r} is the pulse repetition period, d is the antenna element spacing, and $\lceil \cdot \rceil$ rounds a real-valued number towards infinity.

Likewise, we can approximate the rank of the disturbance covariance matrix for the PMF detector as

$$r_{\text{c}} \approx \lceil J + (P-1)\beta \rceil. \tag{3.39}$$

The smaller rank r_{c} over (3.38) is due to the fact that the disturbance covariance matrix is formed over P pulses, which is sufficient for the reduced-dimensional PMF detector due to the underlying AR(P) model. Meanwhile, the space-time disturbance covariance matrix for the full-dimensional MF detector is formed over N (the entire number of) pulses. As such, the PMF can benefit more from the fast convergence property of the CG algorithm.

3.4.5 Preconditioned Conjugate-Gradient PMF

In cases where the disturbance covariance does not have a low-rank structure as in (3.37), preconditioning is usually helpful in improving the convergence rate. The idea is based on the fact that

the convergence rate of CG is determined mainly by the eigenvalue structure of $\mathbf{R_y}$. In particular, the residue between the Wiener solution and kth-step conjugate gradient result is bounded by [10]

$$\|\mathbf{B}_{k,j} - \mathbf{B}_j\|_{\mathbf{R_y}} \leq 2\|\mathbf{B}_{0,j} - \mathbf{B}_j\|_{\mathbf{R_y}} \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k \tag{3.40}$$

where $\|\boldsymbol{\omega}\|_{\mathbf{R_y}} = \sqrt{\boldsymbol{\omega}^H \mathbf{R_y} \boldsymbol{\omega}}$ denotes the $\mathbf{R_y}$ norm and κ is the condition number of $\mathbf{R_y}$. It is clear that rapid convergence can be achieved if κ is near 1. In the following, we discuss the use of preconditioning with the CG-PMF. For simplicity, the resulting detector is referred to as the PCG-PMF detector.

Specifically, consider the modified Wiener-Hopf equation (cf. (3.28))

$$\tilde{\mathbf{R}}_{\mathbf{y}} \tilde{\mathbf{B}}_j = \tilde{\mathbf{R}}_j \tag{3.41}$$

where $\tilde{\mathbf{R}}_{\mathbf{y}} = \mathbf{M}^{-\frac{1}{2}} \mathbf{R_y} \mathbf{M}^{-\frac{1}{2}}$, $\tilde{\mathbf{B}}_j = \mathbf{M}^{\frac{1}{2}} \mathbf{B}_j$, $\tilde{\mathbf{R}}_j = \mathbf{M}^{-\frac{1}{2}} \mathbf{B}_j$, and $\mathbf{M}$ is a Hermitian positive-definite matrix that is called preconditioner [10]. The preconditioner is used to yield a better conditioned $\tilde{\mathbf{R}}_{\mathbf{y}}$, which has a smaller condition number than $\mathbf{R_y}$, and thus a faster convergence rate. For PMF, the disturbance covariance matrix is a block-Toeplitz (BT) matrix. For such matrices, block-circulant (BC) preconditioners are often recommended [22, 23]. Our BC preconditioner can be directly computed from the disturbance covariance matrix $\mathbf{R_y}$ which has the following block Toeplitz matrix structure:

$$\mathbf{R_y} = \begin{bmatrix} \mathbf{R_x}(0) & \cdots & \mathbf{R_x}(P-1) \\ \vdots & \ddots & \vdots \\ \mathbf{R_x}(1-P) & \cdots & \mathbf{R_x}(0) \end{bmatrix} \tag{3.42}$$

where $\mathbf{R_x}(m) = E[\mathbf{x}(n)\mathbf{x}^H(n-m)] \in \mathbb{C}^{J \times J}$. In particular, the BC preconditioner is given by [27]

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_0 & \mathbf{M}_{P-1} & \cdots & \mathbf{M}_1 \\ \mathbf{M}_1 & \mathbf{M}_0 & \cdots & \mathbf{M}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{M}_{P-1} & \mathbf{M}_{P-2} & \cdots & \mathbf{M}_0 \end{bmatrix}$$

where

$$\mathbf{M}_k = \frac{(P-k)\mathbf{R_x}(k) + k\mathbf{R_x}(k-P)}{P}, \qquad 0 \leq k < P. \tag{3.43}$$

It is noted that, as shown in [10], practically $\mathbf{M}^{-\frac{1}{2}}$ does not need to be explicitly calculated in the PCG algorithm. The PCG algorithm is summarized as follows.

Initialization. Initialize the conjugate-direction vector $\mathbf{D}_{1,j}$, gradient vector $\boldsymbol{\gamma}_{1,j}$, preconditioned vector $\mathbf{z}_{1,j}$ and initial solution $\mathbf{B}_{0,j}$:

$$\boldsymbol{\gamma}_{1,j} = -\mathbf{R}_j \tag{3.44}$$

$$\mathbf{D}_{1,j} = \mathbf{z}_{1,j} = \mathbf{M}^{-1}\boldsymbol{\gamma}_{1,j} \tag{3.45}$$

$$\mathbf{B}_{0,j} = \mathbf{0}. \tag{3.46}$$

for $k = 1, 2, \cdots$, till convergence ($k \leq J(P-1)$) **do**

Update the step size $\alpha_{k,j}$:

$$\alpha_{k,j} = \frac{\boldsymbol{\gamma}_{k,j}^H \mathbf{z}_{k,j}}{\mathbf{D}_{k,j}^H \mathbf{R}_{\mathbf{y}} \mathbf{D}_{k,j}}. \tag{3.47}$$

Update the solution $\mathbf{B}_{k,j}$:

$$\mathbf{B}_{k,j} = \mathbf{B}_{k-1,j} + \alpha_{k,j}\mathbf{D}_{k,j}. \tag{3.48}$$

Update the gradient vector $\boldsymbol{\gamma}_{k+1,j}$:

$$\boldsymbol{\gamma}_{k+1,j} = \boldsymbol{\gamma}_{k,j} + \alpha_{k,j}\mathbf{R}_{\mathbf{y}}\mathbf{D}_{k,j}. \tag{3.49}$$

Update the preconditioned vector $\mathbf{z}_{k+1,j}$:

$$\mathbf{z}_{k+1,j} = \mathbf{M}^{-1}\boldsymbol{\gamma}_{k+1,j}. \tag{3.50}$$

Update the conjugate-direction vector $\mathbf{D}_{k+1,j}$

$$\mathbf{D}_{k+1,j} = \mathbf{z}_{k,j} + \frac{\boldsymbol{\gamma}_{k+1,j}^H \mathbf{z}_{k+1,j}}{\boldsymbol{\gamma}_{k,j}^H \mathbf{z}_{k,j}} \mathbf{D}_{k,j}. \tag{3.51}$$

end for

The complexity associated with the AR parameter estimation in PCG-PMF is summarized in TABLE I, where r is the number of iterations needed by the PCG algorithm to reach convergence, and the flop counts are for all J channels. It is interesting to note that the PCG-PMF is computationally very efficient, involving approximately $\mathcal{O}(rJ^3P\log_2 P)$. The computational efficiency is primarily due to the fast convergence rate offered by preconditioning and the use of a BC preconditioner, as explained next. In the following, we discuss the complexity of only $\mathbf{M}^{-1}$, (3.47) and (3.50), since the other calculations are obvious.

First, we consider $\mathbf{M}^{-1}$. Since $\mathbf{M}$ is a block-circulant (BC) matrix, the inverse of $\mathbf{M}$ can be computed by using the Fast Fourier transform (FFT) [28]

$$\mathbf{M}^{-1} = \begin{bmatrix} \boldsymbol{C}_0 & \boldsymbol{C}_{P-1} & \cdots & \boldsymbol{C}_1 \\ \boldsymbol{C}_1 & \boldsymbol{C}_0 & \cdots & \boldsymbol{C}_2 \\ \vdots & \vdots & \cdots & \vdots \\ \boldsymbol{C}_{P-1} & \boldsymbol{C}_{P-2} & \cdots & \boldsymbol{C}_0 \end{bmatrix} \tag{3.52}$$

Table 3.1: Complexity of PCG-PMF

Equation	Flops	Remark
(3.43)	$\mathcal{O}(J^2P)$	calculated once
$\mathbf{M}^{-1}$	$\mathcal{O}(J^2P\log_2 P + J^3P)$	calculated once
(3.47)	$\mathcal{O}(J^3P\log_2 P)$	at k-th iteration
(3.48)	$\mathcal{O}(J^2P)$	at k-th iteration
(3.49)	$\mathcal{O}(J^2P)$	at k-th iteration
(3.50)	$\mathcal{O}(J^3P\log_2 P)$	at k-th iteration
(3.51)	$\mathcal{O}(J^2P)$	at k-th iteration
Total	$\approx \mathcal{O}(rJ^3P\log_2 P)$	for r iterations

where

$$\boldsymbol{C}_m = \frac{1}{P^2}\sum_{k=0}^{P-1} W_P^{-km}\mathbf{M}_k^{-1}, m = 0, 1, \cdots, P-1 \tag{3.53}$$

and $W_P^{-km} = \exp(j2km\pi/P)$. It follows that the computation of $\mathbf{M}^{-1}$ is composed of P matrix inversions of $J \times J$ matrices and J^2 FFTs of length P. Therefore, the total complexity is $\mathcal{O}(J^2P\log_2 P + J^3P)$.

Second, we consider (3.47). The main complexity of (3.47) is matrix-vector multiplication $\mathbf{R}_\mathbf{y}\mathbf{D}_{k,j}$. Since $\mathbf{R}_\mathbf{y}$ is a JP-dimensional BT matrix, the above matrix-vector multiplication consists of J^2 Toeplitz matrix-vector multiplications, where each Toeplitz matrix is a $P \times P$ matrix. Each Toeplitz matrix-vector multiplication can be implemented by the FFT using $\mathcal{O}(P\log_2 P)$ flops [29]. Hence, the complexity of (3.47) for each channel per iteration is $\mathcal{O}(J^2P\log_2 P)$. With J channels and r iterations, the total complexity of (3.47) is $\mathcal{O}(rJ^3P\log_2 P)$.

Finally, we consider (3.50). Since the preconditioner $\mathbf{M}$ is a BC matrix, (3.50) can again be computed by J^2 FFTs of length P. The complexity for each channel per iteration is $\mathcal{O}(J^2P\log_2 P)$, so the total complexity of (3.50) for J channels is $\mathcal{O}(rJ^3P\log_2 P)$.

Here, we make a comparison between the PCG-PMF and CG-PMF. Since the condition number of the preconditioned disturbance covariance matrix $\tilde{\mathbf{R}}_\mathbf{y}$ is generally smaller than that of $\mathbf{R}_\mathbf{y}$, PCG-PMF provides a faster convergence than CG-PMF. The latter has a complexity of $\mathcal{O}(J^3P^3)$.

3.5 Conjugate-Gradient PAMF

The CG-PMF algorithm is now extended to the adaptive case when both the covariance matrix and the AR model order P are unknown. The resulting detector is referred to as the CG-PAMF detector. The extension of CG-PMF involves i) replacing the true covariance matrices with estimates obtained from the target-free training data; and ii) integrating the CG-based model order selection proposed in Section 3.4.3 with conjugate-gradient iterations. The CG-PAMF detector with order selection (OSCG-PAMF) is summarized next.

- ***Step 1:*** Estimate the disturbance covariance matrices from the training data via temporal

and range averaging:

$$\hat{\mathbf{R}}_{\mathbf{y}}^{(\bar{P})} = \begin{bmatrix} \hat{\mathbf{R}}_{\mathbf{x}}(0) & \cdots & \hat{\mathbf{R}}_{\mathbf{x}}(\bar{P}-1) \\ \vdots & \ddots & \vdots \\ \hat{\mathbf{R}}_{\mathbf{x}}(1-\bar{P}) & \cdots & \hat{\mathbf{R}}_{\mathbf{x}}(0) \end{bmatrix} \tag{3.54}$$

$$\hat{\mathbf{R}}_{\mathbf{yx}}^{(\bar{P})} = \begin{bmatrix} \hat{\mathbf{R}}_{\mathbf{x}}(-1) \\ \vdots \\ \hat{\mathbf{R}}_{\mathbf{x}}(-\bar{P}) \end{bmatrix} \tag{3.55}$$

where the sub-matrices are given by

$$\hat{\mathbf{R}}_{\mathbf{xx}}^{H}(m) = \frac{1}{NK} \sum_{k=1}^{K} \sum_{l=1}^{N-m} \mathbf{x}_k(l+m)\mathbf{x}_k^H(l) \tag{3.56}$$

with K denoting the number of training data vectors and $\bar{P}$ determined by (3.35).

- ***Step 2:*** Use the CG algorithm to solve

$$\hat{\mathbf{R}}_{\mathbf{y}}^{(\bar{P})}\hat{\mathbf{B}}^{(\bar{P})} = \hat{\mathbf{R}}_{\mathbf{yx}}^{(\bar{P})}. \tag{3.57}$$

  - ***Step 2.1:*** Examine the residual $\bar{\bar{\epsilon}}_{Jp}$ (3.36) at each Jp-th ($p = 1, 2, \cdots, \bar{P}$) iteration of CG. If $\bar{\bar{\epsilon}}_{Jp} < \alpha_0 \bar{\bar{\epsilon}}_{J(p-1)}$, where $0 < \alpha_0 < 1$ is a stopping threshold, then exit the CG iteration, and set the AR model order as $\hat{P} = p$.

Unlike the original PAMF with an unknown AR model order [16], which has to run recursively from $p = 1$ to a $\hat{P}$ ($\hat{P} \leq \bar{P}$) to jointly estimate the AR coefficients and model order, OSCG-PAMF does not contain any recursion. It only has to perform CG with the disturbance covariance matrix $\mathbf{R}_{\mathbf{y}}^{(\bar{P})}$ for $J\hat{P}$ iterations to obtain a model order estimate.

Remark: Several estimators can be employed to find the linear prediction filters for the PAMF. The estimator as represented by (3.57) along with the covariance matrix estimates (3.54)-(3.56) is often called the multichannel Yule-Walker method. Other estimators, such as the least-squares estimators [16], solve slightly modified versions of the linear equation (3.57). It is noted that in most cases, the CG algorithm can be used to efficiently solve such a modified linear equation. Due to space limit, we will not explore these alternative CG-PAMF detectors.

A similar comparison can be made between the complexity of the OSCG-PAMF detector and that of the eigencanceler when the covariance matrix is unknown. In addition to the numbers of flops as summarized in Section 3.4.3, both have to pay the extra complexity needed to estimate the covariance matrix. In this case, the OSCG-PAMF requires an additional complexity of $\mathcal{O}(J^2\bar{P}NK)$ as incurred in (3.54)–(3.56), whereas the extra complexity for the eigencanceler is $\mathcal{O}(J^2N^2K)$ that is used to estimate a full $(JN \times JN)$ space-time covariance matrix from K training signals.

3.6 Numerical Results

In this section, simulation results are provided to illustrate the performance of the proposed techniques. We consider simulated data generated using AR models and the KASSPER data [20], which were obtained from more realistic clutter models. The simulation results presented below use 20000 independent Monte Carlo data realizations and a probability of false alarm of $P_{\text{fa}} = 10^{-2}$. The chosen P_{fa} may be considered too high for many practical detection applications. It is noted that the choice is only to facilitate computer simulation and reduce simulation time. The main observations from the simulation, including the convergence of the CG algorithm in PAMF detection and the accuracy of the estimated P provided by the proposed model order selection method, are independent of the choice of P_{fa}.

A major issue that we like to illustrate in the following numerical examples is the convergence of the CG algorithm, with partial or full iterations, when the data covariance matrix is known or estimated, and/or when the AR model order is known or estimated. To this end, we compare the various detectors used with the CG algorithm, including the CG-PMF (Section 3.4.2), CG-PAMF with a known AR model order (Section 3.5) and OSCG-PAMF with an estimated model order (Section 3.5), with the same detectors used with direct matrix inverse (DMI). For example, the DMI-PAMF detector involves a direct inverse of the estimated covariance matrix in (3.54) and uses it compute the linear prediction filter (3.28). This DMI approach turns out to coincide with the Yule-Walker method [26] for AR spectral estimation. It is noted in [16] that there are alternative spectral estimation methods which may yield better detection performance in some scenarios. These alternatives are not considered here since the focus is the convergence of the CG algorithm in PAMF. In the following, we will primarily use, as a comparison metric, the probability of detection versus the SINR for a given probability of false alarm. The output SINR of the PAMF detector was derived and extensively studied in [30].

First, we examine the performance of the two implementations of the PMF detector by using simulated data with AR disturbances. The disturbance is an AR(2) process with $J = 4$ elements and $N = 64$ pulses. Both PMF detectors have knowledge of the exact disturbance covariance matrix; however, they use different approaches to compute the linear predictor. Specifically, we consider the DMI-PMF, which uses direct matrix inverse to solve the Wiener-Hopf equation, and the CG-PMF as discussed in Section 3.4.2 with knowledge of the AR model order P. The numerical results are shown in Fig. 3.3. It is seen that both implementations yield an identical detection performance.

We next examine the performance of the CG-based AR model order selection method used in the CG-PMF and CG-PAMF detectors with an unknown AR model order. Two AR disturbance signals with $J = 4$ and $N = 64$ are considered, and their model orders are $P = 1$ and 3, respectively. We choose the same upper bound $\bar{P} = 6$ for both cases. The residual $\bar{\hat{\epsilon}}_k$ (3.36) is computed and used for model order selection; as a benchmark, we also include $\bar{\epsilon}_k$, which is similarly computed as in (3.36) but with $\hat{\boldsymbol{\epsilon}}_{k,j}$ replaced by $\boldsymbol{\epsilon}_{k,j}$. Recall that $\bar{\hat{\epsilon}}_k$ is an approximation of $\bar{\epsilon}_k$, which cannot be computed in practice due to the fact that the true Wiener-Hopf solution is unknown. The numerical results for CG-PMF are shown in Fig. 3.4 and Fig. 3.5, which correspond to $P = 1$ and $P = 3$, respectively. It is seen that, $\bar{\hat{\epsilon}}_k$ has a sharp decrease at the JP-th

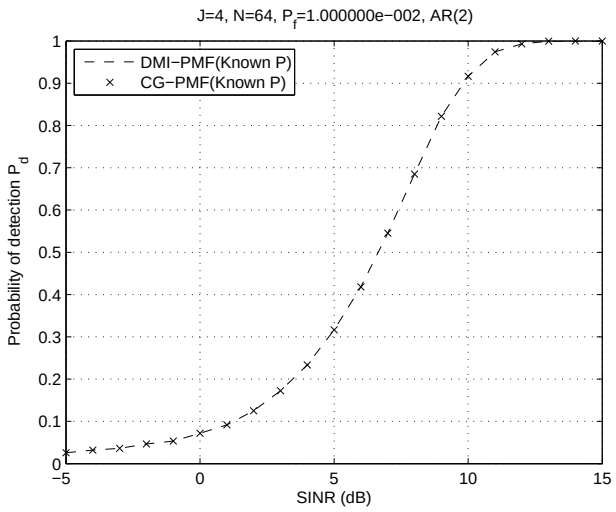

Figure 3.3: Probability of detection versus SINR of PMF for simulated data($J = 4$; $N = 64$; $P = 2$)

($JP = 4$ for $P = 1$ and $JP = 12$ for $P = 3$) iteration of CG, which confirms the effectiveness of the CG-based model order selection method. The counterpart model order selection results for CG-PAMF are shown in Fig. 3.6 and Fig. 3.7, for $P = 1$ and $P = 3$, respectively. Unlike the CG-PMF which uses the real disturbance covariance matrix, the sample covariance matrix estimated from the training data (cf. (3.42)) is employed for model order selection in CG-PAMF. Here the training data size is set to $K = 32$. It is also shown that $\bar{\tilde{\epsilon}}_k$ has a sharp decrease at the JP-th ($JP = 4$ for $P = 1$ and $JP = 12$ for $P = 3$) iteration of CG, although the decrease in residue is smaller than that of CG-PMF due to estimation error of the sample covariance matrix.

We now consider the convergence of PCG-PMF. The simulated disturbance is an AR(8) multichannel process with $J = 4$. The convergence of CG-PMF and PCG-PMF is shown in Fig. 3.8. The condition number of the preconditioned covariance matrix is 4.2, which is much less than the condition number of the original covariance matrix 77.1. It is seen from Fig. 3.8 that only 5 iterations are needed in PCG-PMF to reach a relative approximation error under 1%, while 20 iterations are needed for CG-PMF.

Our next example considers the adaptive PAMF detector, for which the disturbance covariance matrix is unknown and the sample covariance matrix is estimated by (3.54). Similar to the PMF detector, we compare two implementations of the PAMF detector, including the DMI-PAMF and CG-PAMF, Here the DMI-PAMF directly inverses the sample covariance matrix to get the maximum-likelihood estimation of AR coefficients [26]. The disturbance is an AR(2) signal, whose disturbance covariance matrix is estimated from $K = 16$ target-free training data vectors, and the AR coefficients are estimated based on the estimated disturbance covariance matrix. The numerical results are shown in Fig. 3.9. It is observed that both implementations yield an identical detection performance.

The performance of the CG-PAMF with an unknown disturbance AR model order and dis-

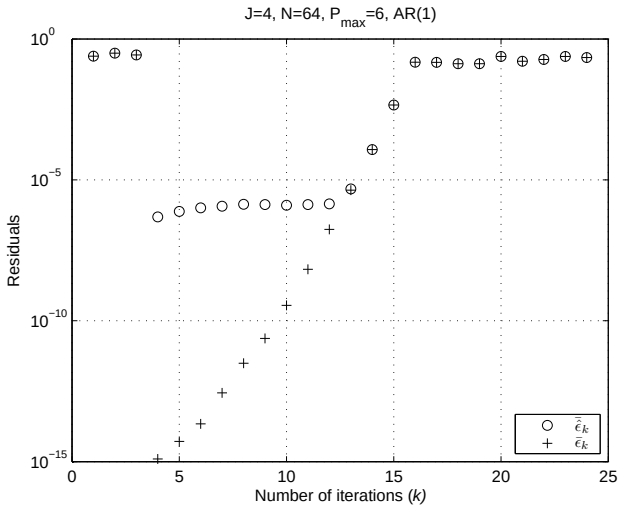

Figure 3.4: Residuals $\bar{\epsilon}_{Jp}$ and $\bar{\hat{\epsilon}}_{Jp}$ for model order selection in CG-PMF ($J = 4$; $N = 64$; $P = 1$; $\bar{P} = 6$)

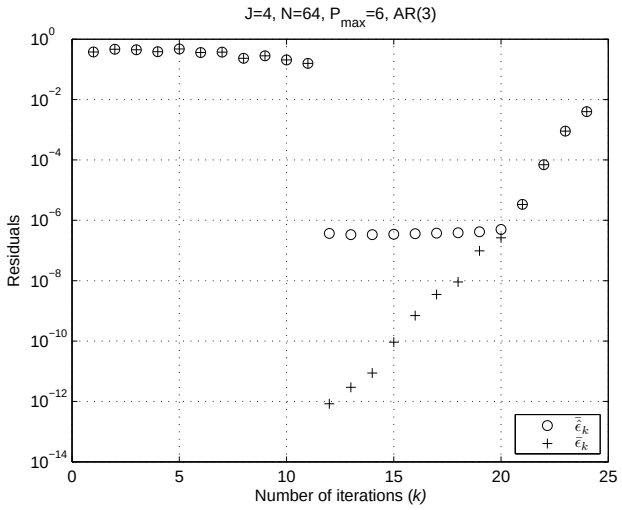

Figure 3.5: Residuals $\bar{\epsilon}_{Jp}$ and $\bar{\hat{\epsilon}}_{Jp}$ for model order selection in CG-PMF ($J = 4$; $N = 64$; $P = 3$; $\bar{P} = 6$)

turbance covariance matrix is considered next. Both AR model based data and KASSPER 2002 data set are employed in this example. The KASSPER data set was generated by considering practical airborne radar parameters and issues found in a real-world clutter environment [20]. Specifically, the simulated airborne radar platform travels at a speed of 100 m/s with a 3° crab angle. The radar carrier frequency is 1240 MHz. The horizontal 11 antenna elements form a ULA with a spacing of 0.1092m between adjacent elements, and the transmit array is uniformly weighted and phased to steer the main beam to 195°. The pulse repetition frequency is 1984 Hz and a coherent processing interval contains 32 pulses. Only the first 8 elements are used in our

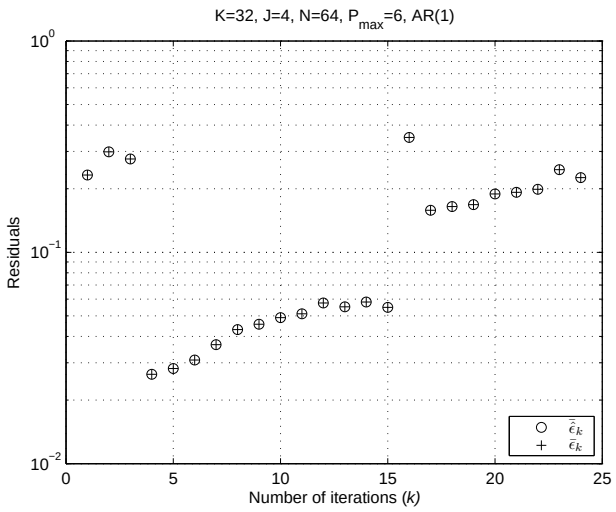

Figure 3.6: Residuals $\bar{\epsilon}_{Jp}$ and $\bar{\hat{\epsilon}}_{Jp}$ for model order selection in CG-PAMF ($K = 32$; $J = 4$; $N = 64$; $P = 1$; $\bar{P} = 6$)

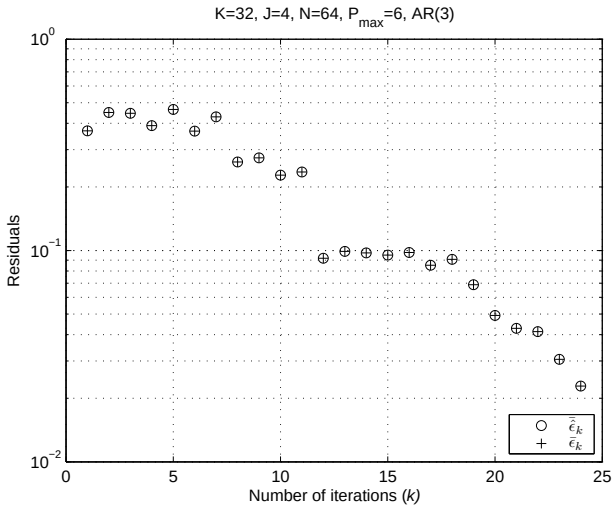

Figure 3.7: Residuals $\bar{\epsilon}_{Jp}$ and $\bar{\hat{\epsilon}}_{Jp}$ for model order selection in CG-PAMF ($K = 32$; $J = 4$; $N = 64$; $P = 3$; $\bar{P} = 6$)

simulation. We use the covariance matrix associated with range bin 200 in the KASSPER data set to generate the test data and the covariance matrices from the neighboring ranges bins to generate the training signals. A target is injected into the test cell with a normalized spatial frequency 0.1 and a normalized Doppler frequency 0.35.

The numerical results are shown in Fig. 3.10 for the AR model based data and, respectively, Fig. 3.11 for the KASSPER 2000 data, where OSCG-PAMF (unknown P) represents the CG-PAMF detector with the CG-based model order selection method which employs a model order

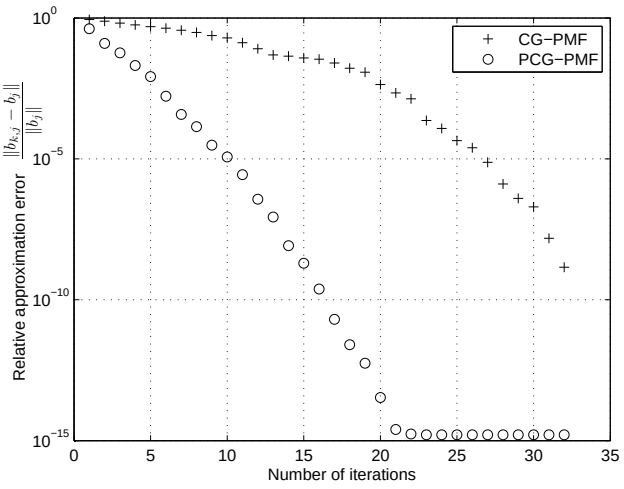

Figure 3.8: Convergence of CG-PMF and PCG-PMF ($J = 4$; $P = 8$)

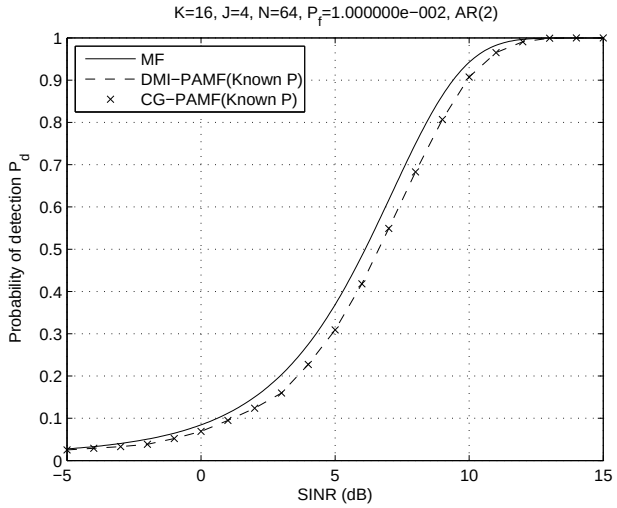

Figure 3.9: Probability of detection versus SINR of PAMF for AR data ($K = 16$; $J = 4$; $N = 64$; $P = 2$)

upper bound $\bar{P}$ calculated by (3.35). In Fig. 3.11, we also include for comparison the joint domain localized (JDL) detector [31], a popular reduced-dimensional STAP solution in scenarios of limited training. The JDL is implemented by using 3 beams and 3 Doppler bins for adaptivity. It is seen that the performance of the OSCG-PAMF is nearly identical to that of CG-PAMF with known P (AR data) or a pre-selected $P = 2$ (KASSPER data). For the case of AR data, we noticed that only one model order selection error ($\hat{P} \neq P$) occurred out of 20000 simulations. Moreover, using the relevant parameters of the KASSPER data, we have $\beta = 2v_{\mathrm{g}}T_{\mathrm{r}}/d = 0.923$. It follows that for $J = 8$ elements, the maximum number of conjugate-gradient iterations needed by the CG-PAMF for a given model order p is estimated to be $r_{cp} + 1 = \lceil 8.077 + 0.923p \rceil$. For

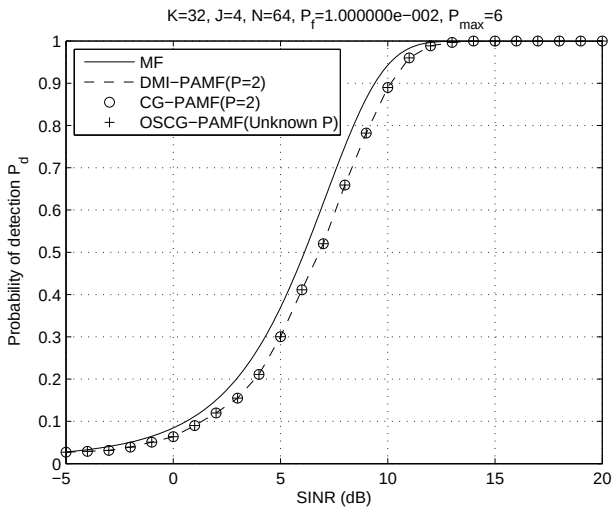

Figure 3.10: Probability of detection versus SINR for AR data ($K = 32$; $J = 4$; $N = 64$; $P = 2$; $\bar{P} = 6$)

example, the maximum numbers of CG iterations for $p = 2$ is about 10 due to the low-rank structure of the clutter, whereas without such a structure, it would require $pJ = 16$ iterations for the CG to converge. It is also seen in Fig. 3.11 that the PAMF detectors outperform the JDL-AMF detector. The JDL-AML experiences a loss of about 4 dB compared with the MF.

Finally, we compare the complexity in terms of the number of flops required by the CG and DMI implementations. The flops required by the CG-PAMF and DMI-PAMF versus the AR model order p are shown in Fig. 3.12. For the DMI-PAMF, the QR decomposition is adopted to get the J-channel AR coefficients. It is seen that the complexity of the CG-PAMF is lower than that of the DMI-PAMF.

3.7 Concluding Remarks

The conjugate-gradient (CG) algorithm was employed to solve the linear prediction problem underlying the parametric matched filter (PMF) and parametric adaptive matched filter (PAMF) detectors. It is shown that the CG algorithm leads to not only new efficient implementations, but also new insights of these parametric detectors as reduced-dimensional subspace detectors. In particular, the linear prediction filter and whitening filter of the PMF and PAMF detectors are within the Krylov subspace of dimension JP, and these detectors are reduced JP-dimensional subspace detectors, where J and P are the number of channels and AR model order, respectively. We examined the convergence rate of the CG parametric detectors. In airborne radar applications, the special low-rank structure of the disturbance covariance matrix implies that a rapid convergence is possible, whereby convergence can be achieved without completing a full round of CG iterations. Even for disturbance covariance matrices that do not have the low-rank structure, preconditioning methods can be used to speed up the convergence rate. In general, the

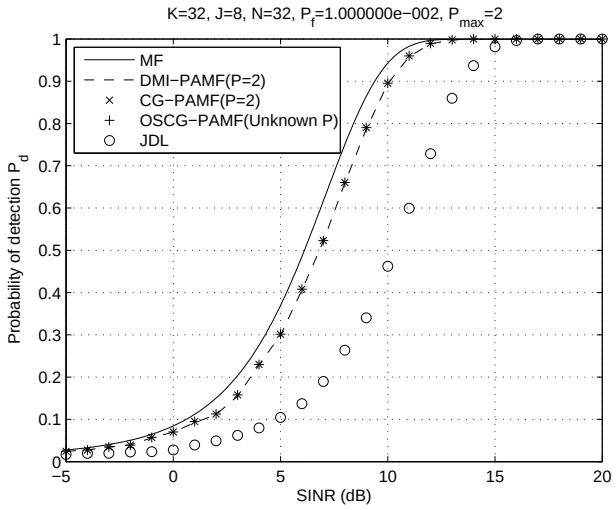

Figure 3.11: Probability of detection versus SINR for KASSPER 2002 data ($K = 32$; $J = 8$; $N = 32$; $P = 2$; $\bar{P} = 2$)

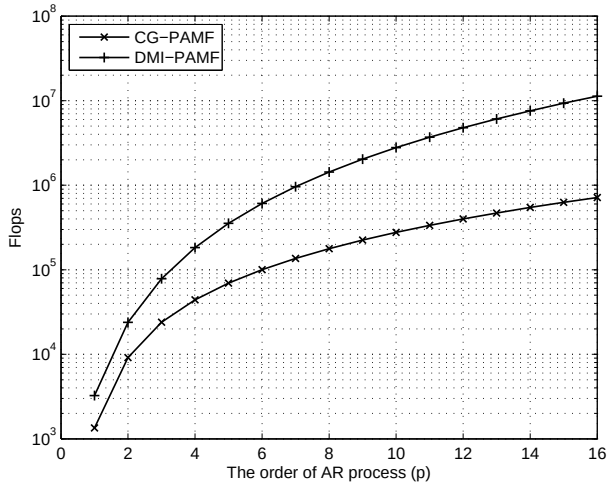

Figure 3.12: Computational complexity of CG-PAMF and DMI-PAMF versus AR model order p ($K = 32$; $J = 8$; $N = 32$).

CG parametric detectors are more efficient than their counterparts implemented in conventional approaches. We also presented a new CG-based AR model order selection method, which is naturally integrated with the CG iterations. The proposed techniques are illustrated by using both KASSPER and other simulated data.

Finally, we note that the CG algorithm bears some similarity to a vector space approach [32] to solving the multi-dimensional Yule-Walker equation for an arbitrary region of support. Both involve the use of conjugate orthogonal basis vectors. A future subject would be to investigate

the relation of the two approaches and explore the application of the CG algorithm for multi-dimensional and multichannel applications.

3.8 Appendix: Proof of Lemma 1

It is well known that the conjugate direction vectors obtained by the CG algorithm solving the Wiener-Hopf equation (3.30) span the Krylov subspace [10]:

$$\mathcal{K}(\mathbf{R}_j^{(\bar{P})}, \mathbf{R}_{\mathbf{y}}^{(\bar{P})}, k) = \text{span}\left\{\mathbf{D}_{1,j}^{(\bar{P})}, \mathbf{D}_{2,j}^{(\bar{P})}, \cdots, \mathbf{D}_{k,j}^{(\bar{P})}\right\} \tag{3.58}$$

Furthermore, the truncated solution obtained at the k-th iteration $\mathbf{B}_{k,j}^{(\bar{P})}$ minimizes the $\mathbf{R}_{\mathbf{y}}^{(\bar{P})}$-norm of the approximation error over all vectors on $\mathcal{K}(\mathbf{R}_j, \mathbf{R}_{\mathbf{y}}^{(\bar{P})}, k)$ [33], i.e.

$$\left\|\mathbf{B}_j^{(\bar{P})} - \mathbf{B}_{k,j}^{(\bar{P})}\right\|_{\mathbf{R}_{\mathbf{y}}^{(\bar{P})}} = \min_{a_k} \left\|\mathbf{x} - \sum_{k=0}^{r} a_k \mathbf{R}_{\mathbf{y}}^{(\bar{P})\,k} \mathbf{B}_j^{(\bar{P})}\right\|_{\mathbf{R}_{\mathbf{y}}^{(\bar{P})}}. \tag{3.59}$$

Therefore, the truncated solution obtained at the k-th iteration $\mathbf{B}_{k,j}^{(\bar{P})}$ is the $\mathbf{R}_{\mathbf{y}}^{(\bar{P})}$-orthogonal projection of the Wiener solution $\mathbf{B}_j^{(\bar{P})}$ to the subspace $\mathcal{K}(\mathbf{R}_j, \mathbf{R}_{\mathbf{y}}^{(\bar{P})}, k)$, and

$$\boldsymbol{\alpha}_{k,j} = \left[\alpha_{1,j}^{(\bar{P})}, \alpha_{2,j}^{(\bar{P})}, \cdots, \alpha_{k,j}^{(\bar{P})}\right]^T \tag{3.60}$$

contains the coordinate values of conjugate-direction vectors $\left\{\mathbf{D}_{1,j}^{(\bar{P})}, \mathbf{D}_{2,j}^{(\bar{P})}, \cdots, \mathbf{D}_{k,j}^{(\bar{P})}\right\}$, which are given by

$$\alpha_{k,j}^{(\bar{P})} = \frac{\mathbf{D}_{k,j}^{(\bar{P})\,H} \mathbf{R}_{\mathbf{y}}^{(\bar{P})} \mathbf{B}_j^{(\bar{P})}}{\mathbf{D}_{k,j}^{(\bar{P})\,H} \mathbf{R}_{\mathbf{y}}^{(\bar{P})} \mathbf{D}_{k,j}^{(\bar{P})}}. \tag{3.61}$$

With the definition of $\tilde{\mathbf{D}}_{k,j}$ by (3.32), we can write after JP iterations

$$\boldsymbol{\alpha}_{JP,j} = \tilde{\mathbf{D}}_{JP,j}^H \mathbf{B}_j^{(\bar{P})} \tag{3.62}$$

where $\tilde{\mathbf{D}}_{JP,j} = [\tilde{\mathbf{D}}_{1,j}, \tilde{\mathbf{D}}_{2,j}, \cdots, \tilde{\mathbf{D}}_{JP,j}]$. Recalling $\mathbf{B}_j^{(\bar{P})} = [\mathbf{B}_j^T, \mathbf{0}_{J(\bar{P}-P)\times 1}^T]^T$, we have

$$\boldsymbol{\alpha}_{JP,j} = \begin{bmatrix}\bar{\mathbf{D}}_{JP,j}^H & \tilde{\mathbf{D}}_{JP,\mathrm{d}}^H\end{bmatrix} \begin{bmatrix}\mathbf{B}_j \\ \mathbf{0}\end{bmatrix} \tag{3.63}$$

where $\bar{\mathbf{D}}_{JP,j} \in \mathbb{C}^{JP\times JP}$ is the upper $JP \times JP$ block matrix of $\tilde{\mathbf{D}}_{JP,j}$, and $\tilde{\mathbf{D}}_{JP,\mathrm{d}}$ contains the lower block of $\tilde{\mathbf{D}}_{JP,j}$. Then $\mathbf{B}_j$ is given by

$$\mathbf{B}_j = \bar{\mathbf{D}}_{JP,j}^{-H} \boldsymbol{\alpha}_{JP,j}. \tag{3.64}$$

The intermediate solution obtained at the JP-th CG iteration is

$$\mathbf{B}_{JP,j}^{(\bar{P})} = \sum_{m=1}^{JP} \alpha_{m,j} \mathbf{D}_{m,j}^{(\bar{P})} = \mathbf{D}_{JP,j}^{(\bar{P})} \boldsymbol{\alpha}_{JP,j}. \tag{3.65}$$

It follows from (3.64) and (3.65) that $\mathbf{B}_{JP,j}^{(\bar{P})}$ and $\mathbf{B}_j$ are related by

$$\mathbf{B}_{JP,j}^{(\bar{P})} = \mathbf{D}_{JP,j}^{(\bar{P})} \bar{\mathbf{D}}_{JP,j}^{H} \mathbf{B}_j = \mathbf{W}_{JP,j} \mathbf{B}_j \tag{3.66}$$

where $\mathbf{W}_{JP,j} = \mathbf{D}_{JP,j}^{(\bar{P})} \bar{\mathbf{D}}_{JP,j}^{H} \in \mathbb{C}^{J\bar{P} \times JP}$, which completes the proof.

Bibliography

[1] J. Ward, "Space-time adaptive processing for airborne radar," Lincoln Laboratory, MIT, Technical Report 1015, December 1994.

[2] L. E. Brennan and I. S. Reed, "Theory of adaptive radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 9, no. 2, pp. 237–252, 1973.

[3] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, no. 1, pp. 208–216, January 1992.

[4] E. J. Kelly, "An adaptive detection algorithm," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 22, no. 1, pp. 115–127, March 1986.

[5] A. Haimovich, "The eigencanceler: Adaptive radar by eigenanalysis methods," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, no. 2, pp. 532–542, April 1996.

[6] I. P. Kirsteins and D. W. Tufts, "Adaptive detection using low rank approximation to a data matrix," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 30, no. 1, pp. 55–67, January 1994.

[7] J. S. Goldstein and I. S. Reed, "Reduced-rank adaptive filtering," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 492–496, February 1997.

[8] J. S. Goldstein, I. S. Reed, and L. L. Scharf, "A multistage representation of the Wiener filter based on orthogonal projections," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2943–2959, November 1998.

[9] C. D. Peckham, A. M. Haimovich, T. F. Ayoub, J. S. Goldstein, and I. S. Reed, "Reduced-rank STAP performance analysis," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, no. 2, pp. 664–676, April 2000.

[10] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins University Press, 1996.

[11] M. E. Weippert, J. D. Hiemstra, J. S. Goldstein, and M. D. Zoltowski, "Insights from the relationship between the multistage Wiener filter and the method of conjugate gradients," in

Sensor Array and Multichannel Signal Processing Workshop Proceedings, 2002, Rosslyn, VA, August 2002, pp. 388–392.

[12] L. L. Scharf, E. K. P. Chong, M. D. Zoltowski, J. S. Goldstein, and I. S. Reed, "Subspace expansion and the equivalence of conjugate direction and multistage wiener filters," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 5013–5019, October 2008.

[13] H. Ge, I. R. Kirsteins, and L. L. Scharf, "Data dimension reduction using Krylov subspaces: Making adaptive beamformers robust to model order-determination," in *Proceedings of the 2006 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 4, Toulouse, France, May 2006, pp. 1001–1004.

[14] I. R. Kirsteins and H. Ge, "Performance analysis of Krylov space adaptive beamformers," in *Proceedings of the Fourth IEEE Workshop on Sensor Array and Multichannel Processing (SAM)*, Waltham, MA, July 2006, pp. 16–20.

[15] C. Jiang, H. Li, and M. Rangaswamy, "Conjugate gradient parametric detection of multichannel signals," *IEEE Transactions on Aerospace and Electronic Systems*, to appear.

[16] J. R. Román, M. Rangaswamy, D. W. Davis, Q. Zhang, B. Himed, and J. H. Michels, "Parametric adaptive matched filter for airborne radar applications," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, no. 2, pp. 677–692, April 2000.

[17] K. J. Sohn, H. Li, and B. Himed, "Parametric Rao test for multichannel adaptive signal detection," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 3, pp. 920–933, July 2007.

[18] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE Signal Processing Magazine*, pp. 4–24, April 1988.

[19] M. Rangaswamy, F. C. Lin, and K. R. Gerlach, "Robust adaptive signal processing methods for heterogeneous radar clutter scenarios," *Signal Processing*, vol. 84, no. 9, pp. 1477–1733, September 2004.

[20] J. S. Bergin and P. M. Techau, "High-fidelity site-specific radar simulation: KASSPER'02 workshop datacube," Information Systems Laboratories, Inc., Vienna, VA, Technical Report ISL-SCRD-TR-02-105, May 2002.

[21] M. Rangaswamy and J. H. Michels, "A parametric multichannel detection algorithm for correlated non-Gaussian random processes," in *Proceedings of the 1997 IEEE National Radar Conference*, Syracuse, NY, May 1997, pp. 349–354.

[22] T. Chan and J. Olkin, "Circulant preconditioners for Toeplitz-block matrices," *Numerical Algorithmns.*, vol. 6, no. 1, pp. 89–101, March 1994.

[23] X.-Q. Jin, *Developments and Applications of Block Toeplitz Iterative Solvers*. Science Press, Kluwer Academic Publishers, 2002.

[24] I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 10, no. 6, pp. 853–863, 1974.

[25] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*. New York, NY: John Wiley & Sons, 1968.

[26] S. M. Kay, *Modern Spectral Estimation: Theory and Application*. Englewood Cliffs, NJ: Prentice Hall, 1988.

[27] T. Chan, "An optimal circulant preconditioner for Toeplitz systems," *SIAM Journal on Scientific and Statistical Computing.*, vol. 9, pp. 766–771, 1988.

[28] R. Vescovo, "Inversion of block-circulant matrices and circular array approach," *IEEE Transactions on Antennas and Propagation*, vol. 45, no. 10, pp. 1565–1567, October 1997.

[29] R. H. Chan and M. K. Ng, "Conjugate gradient methods for Toeplitz systems," *SIAM Review*, vol. 38, no. 3, pp. 427–482, September 1996.

[30] J. H. Michels, J. R. Roman, and B. Himed, "Beam control using the parametric adaptive matched filter stap approach," in *Proceedings of 2003 IEEE Radar Conference*, Huntsville, AL, May 2003, pp. 405–412.

[31] H. Wang and L. Cai, "On adaptive spatial-temporal processing for airborne surveillance radar systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 30, no. 3, pp. 660–670, July 1994.

[32] S. Kay and C. Carbone, "Vector space solution to the multidimensional Yule-Walker equations," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, Hong Kong, China, April 2003, pp. 289–292.

[33] Y. Saad, *Iterative Methods for Sparse Linear Systems*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2003.